ISOPERIMETRIC ESTIMATES ON
SIERPINSKI GASKET TYPE FRACTALS

ROBERT S. STRICHARTZ

ABSTRACT. For a compact Hausdorff space F that is pathwise connected, we can define the *connectivity dimension* β to be the infimum of all b such that all points in F can be connected by a path of Hausdorff dimension at most b. We show how to compute the connectivity dimension for a class of self–similar sets in $\mathbb{R}^n$ that we call *point connected*, meaning roughly that F is generated by an iterated function system acting on a polytope P such that the images of P intersect at single vertices. This class includes the polygaskets, which are obtained from a regular n–gon in the plane by contracting equally to all n vertices, provided n is not divisible by 4. (The Sierpinski gasket corresponds to $n = 3$.) We also provide a separate computation for the octogasket ($n = 8$), which is not point connected. We also show, in these examples, that $\inf \mathcal{H}_\beta(\gamma_{x,y})^{1/\beta}$, where the infimum is taken over all paths $\gamma_{x,y}$ connecting x and y, and $\mathcal{H}_\beta$ denotes Hausdorff measure, is equivalent to the original metric on F.

Given a compact subset F of the plane of Hausdorff dimension α and connectivity dimension β, we can define the *isoperimetric profile function $h(L)$* to be the supremum of $\mathcal{H}_\alpha(F \cap D)$, where D is a region in the plane bounded by a Jordan curve (or union of Jordan curves) γ entirely contained in F, with $\mathcal{H}_\beta(\gamma) \leq L$. The analog of the standard isperimetric estimate is $h(L) \leq cL^{\alpha/\beta}$. We are particularly interested in finding the best constant c and identifying the extremal domains where we have equality. We solve this problem for polygaskets with $n = 3, 5, 6, 8$. In addition, for $n = 5, 6, 8$ we find an entirely different estimate for $h(L)$ as $L \to \infty$, since the boundary of F has infinite $\mathcal{H}_\beta$ measure. We find that the isoperimetric profile function is discontinuous, and that the extremal domains have relatively simple polygonal boundaries. We discuss briefly the properties of minimal paths for the Sierpinski gasket, and the isodiametric problem in the intrinsic metric.

1. INTRODUCTION

Isoperimetric estimates relate the size of a domain to the size of its boundary, and describe the nature of extremal domains for the inequality. Most familiar is the estimate

$$(1.1) \qquad A \leq (4\pi)^{-1} L^2$$

for the area A of a planar domain with perimeter of length L, with the extremal domains being the discs. The delightful survey article of Osserman [O] is a good

introduction to this topic. Note that the power 2 in (1.1) is the ratio of the dimension of the domain to the dimension of the perimeter, and is exactly required in order to have an inequality that is invariant under dilation.

In this paper we obtain analogous estimates for domains in some self–similar fractals F in the plane, such as the Sierpinski gasket. We will only deal with pathwise connected fractals F, and domains bounded by paths entirely contained in F. Suppose such a path Γ is a Jordan curve (or a disjoint union of Jordan curves). Then the interior D of Γ in the plane is well–defined by the Jordan curve theorem, and our domain is just $D \cap F$. We write $V = \mathcal{H}_\alpha(D \cap F)$ for the "volume" of the domain, where α is the Hausdorff dimension of F, and $\mathcal{H}_\alpha$ denotes Hausdorff measure of dimension α. We write $L = \mathcal{H}_\beta(\Gamma)$ for the "length" of the perimeter, for an appropriate minimal choice of β. Then the analogue of (1.1) is

$$(1.2) \qquad\qquad\qquad V \le cL^{\alpha/\beta}.$$

Our goal is to find the best constant c, and to describe the extremal domains for which (1.2) becomes an equality. We would also like, if possible, to compute the isoperimetric profile function $h(L)$, which is defined to be the maximum V value over all domains with $\mathcal{H}_\beta(\Gamma) \le L$, and to describe the nature of such extremal domains.

But first, we need to fix the value of β. To do this we define the *connectivity dimension* β of F to be the infimum of all values b such that any two points in F can be joined by a path of Hausdorff dimension at most b. This definition makes sense for any compact pathwise connected metric space. This definition has been discussed in the physical sciences literature under the name "chemical dimension" [HB], and in the mathematics literature as "exponent of interior distance" [BS], [BK]. Unfortunately, the key paper [BS] has not been published. We always have $1 \le \beta \le \alpha$, since all paths have dimension at least 1, and the dimension of F bounds the dimension of all subsets of F. We are most interested in the case when the regularity estimate

$$(1.3) \qquad\qquad c_1 d(x,y)^\beta \le \mathcal{H}_\beta(\gamma_{x,y}) \le c_2 d(x,y)^\beta$$

holds, where $\gamma_{x,y}$ denotes a minimal (minimizing $\mathcal{H}_\beta(\gamma)$) path connecting x and y. In Section 2 we show how to compute the connectivity dimension for a family of sets we call *point connected self–similar sets*. Some of these results are contained in [BS], but we include the proofs for completeness. These sets are defined to be the attractors of an iterated function system (i.f.s.) $\{S_1, \ldots, S_m\}$ of contractive similarities in $\mathbb{R}^n$ satisfying (i) $S_j P \subseteq P$ for all j, (ii) $\bigcup_{j=1}^m S_j P$ is connected, (iii) if $S_j P$ and $S_k P$ intersect, for $j \ne k$, their intersection consists of a single vertex. Here P denotes the convex hull of the fixed points of $S_1, \ldots, S_m$. We show that the regularity condition (1.3) holds in many cases, such as for the polygaskets which generalize the construction ot the Sierpinski gasket starting with a regular polygon of n sides, provided n is not a multiple of 4. The method of proof actually gives a construction of a family of minimal paths joining the vertices of P, and these paths are also self–similar sets (more precisely, the family is self–similar). Our example of fractals for which we obtain isoperimetric estimates is drawn from this family.

The first example we consider is the standard Sierpinski gasket, in Section 3. The connectivity dimension is obviously 1, and we are able to give a fairly complete description of the isoperimetric profile function, over a portion of its domain,

and to find the best constant in (1.2) (this depends on the normalization of the Hausdorff measure) and the extremal domains. There are two striking differences in comparison with the Euclidean case (1.1). The first is that the isoperimetric profile function is discontinuous; in fact the discontinuities form a countable dense set (since $h(L)$ is monotone increasing, this is the worst possible behavior). The function $h(L)$ is given in terms of a very peculiar function $\psi(x)$, which might be called the base 2 to base 3 conversion function: for x in $[0, 1]$ write $x = .x_1 x_2 \cdots$ base 2, and set $\psi(x) = .x_1 x_2 \cdots$ interpreted as a base 3 number. Obviously $\psi(x)$ has jump discontinuities at all dyadic rational numbers which have ambiguous base 2 representations. The second striking result is that the extremal domains for (1.2) are bounded by polygonal paths. In fact this feature shows up in all the examples we consider. (Whether or not this is always the case is another matter, since we have not studied any examples with nonarithmetic contraction ratios. In related problems [AS] this is a crucial distinction.) This is quite far from the circular boundaries of the extremal domains in the Euclidean case. It is not hard to see that a Sierpinski gasket contains no genuine curves: any C^1 path must be a straight line. The reason for this is that it only contains line segments with 3 distinct slopes, so it is impossible to move continuously among these (this is not a rigorous proof, of course ... in fact we will not use this result in any of our arguments). Thus, if we knew that the extremal domain for (1.2) had to have a piecewise C^1 boundary, then there would be no surprise that the boundary is polygonal. But there seems to be no a priori reason to believe this, and indeed there are many extremal domains for the isoperimetric profile function (for specific values of L) whose boundaries are not polygonal.

In Section 4 we take up the case of the hexagasket. The connectivity dimension is still 1, but the new feature is that there is no boundary of finite length for the whole fractal. Thus we have to let $L \to \infty$ in the domain of the isoperimetric profile function to allow V to approach the maximum value (we normalize $\mathcal{H}_\alpha$ so that $\mathcal{H}_\alpha(F) = 1$ is the maximum). This means that (1.2) becomes uninteresting for large values of L, and we find another type of isoperimetric estimate to describe this asymptotic behavior. It turns out that there is a new exponent, $\gamma = \log(3/2)/\log(4/3)$, such that $1 - V \approx cL^{-\gamma}$ for large domains, and there are corresponding estimates (see Theorem 4.4) of a more complicated nature. Also, it turns out that the extremal domain for (1.2) does not possess the dihedral–6 symmetry of the hexagasket. This is the only example in which the extremal domain is asymmetric, and there seems to be no obvious explanation in either direction. (The principle "symmetric problems have symmetric solutions" is well–known to be notoriously unreliable.)

In Section 5 we study the pentagasket. The connectivity dimension is now greater than 1. The polygonal nature of the extremal domain now must be interpreted as saying the boundary path consists of a finite union of minimal paths. Otherwise, the results follow the same pattern as for the hexagasket.

The octogasket requires special attention, since it is not point–connected. The contracted octagons $S_j P$ intersect along whole edges, and in fact the images $S_j F$ of the fractal have a whole Cantor set of intersection points along these edges. Thus there is a bewilderingly large set of possibilities for the paths in F, and the method of Section 2 does not apply to compute the connectivity dimension. In Section 6 we succeed in computing it, but the method of proof is quite vexing. We give a complicated argument for the existence of complicated minimal paths, and then

observe that there exist much simpler minimal paths. We have not been able to find a direct argument showing that these simpler paths are minimal, but these are the building blocks of the extremal domains that we construct in Section 7. The new feature here is that there is no "big" extremal domain (i.e., one that surrounds the largest bounded component of the complement).

The method of proof we use is very geometric (examining figures). We classify the Jordan curves Γ in terms of the components of the complement of F that they surround, we compute V and L for a family of candidate domains, and we select the extremal domains using elementary calculus and a few numerical computations. Each example requires somewhat different geometric considerations, so we cannot yet present a unified proof.

In Section 8 we discuss briefly some related questions involving the intrinsic metric given in terms of the lengths of minimal paths joining points in F, in the case of the Sierpinski gasket. These minimal paths play the role of geodesics in Riemannian geometry, but their properties are quite different. We discuss also the isodiametric problem in this intrinsic metric.

We note that the isoperimetric problem is not entirely intrinsic. Even though the boundary curve Γ lies entirely in F, the interior of Γ depends on the embedding of F in the plane. Although most fractals are rigid with respect to the inherited metric, this would not be the case using the intrinsic metric. One possible approach would be to start from the domain and try to define its boundary in an intrinsic way. But it is not clear how to do this; the topological boundary is definitely not the appropriate choice.

We also note that there is an entirely different isioperimetric problem: consider an arbitrary Jordan curve in the plane, not necessarily contained in F, of given length L, and maximize the measure of the interior. This problem is even less intrinsic, and has the peculiar feature that the components of the curve in the complement of F would have to be straight line segments. We do not study this problem here.

The Euclidean isoperimetric estimate is closely related to Sobolev inequalities and other analytic estimates [O]. Such analytic estimates are certainly of interest in fractals [M], but there does not seem to be any connection with our isoperimetric estimates.

We assume the reader is familiar with the basic properties of self–similar fractals and Hausdorff measures ([B], [F], [H]). We will use also the generalization to families of self–similar sets ([MW]). The family of polygaskets was studied in [S] in connection with a different question.

I am grateful to Kyallee Dalrymple for creating the images of the pentagasket, hexagasket and octogasket used in the figures.

2. Connectivity dimension

Let F be a compact metric space that is pathwise connected. We define the *connectivity dimension* of F by

(2.1)
$$\beta = \inf\{b : \text{any pair of points in } F \text{ is joined by a path}$$
$$\text{of Hausdorff dimension } \leq b\}.$$

We say that the connectivity dimension is *regular* if there exist constants c_1 and c_2 such that any x, y in F are joined by a path $\gamma_{x,y}$ (minimizing $\mathcal{H}_\beta(\gamma_{x,y})$) with

$$(2.2) \qquad\qquad c_1 d(x,y)^\beta \leq \mathcal{H}_\beta(\gamma_{x,y}) \leq c_2 d(x,y)^\beta,$$

where $\mathcal{H}_\beta$ denotes Hausdorff measure of dimension β. Clearly we always have $\beta \geq 1$, since a path cannot have dimension less than 1, and $\beta \leq \alpha$, where α denotes the Hausdorff dimension of F.

Our goal in this section is to compute the connectivity dimension of a class of fractals related to the Sierpinski gasket, and to show that they are regular. We assume that F is the attractor of an i.f.s. of contractive similarities $\{S_1, \dots, S_m\}$ in $\mathbb{R}^n$. We write

$$(2.3) \qquad\qquad S_j x = \lambda_j + \rho_j R_j (x - \lambda_j),$$

where $\lambda_j \in \mathbb{R}^n$ is the fixed point, $0 < \rho_j < 1$ is the contraction ratio, and R_j is an orthogonal transformation. We refer to F as a *self–similar set*. Let P denote the convex hull of the fixed points $\{\lambda_1, \dots, \lambda_m\}$. P is thus a polytope whose vertices are a subset of the fixed points. Write $P_j = S_j P$, the image polytope under the contraction S_j. We say that F is a *point–connected* self–similar set if the following assumptions hold:

$$(2.4) \qquad\qquad P_j \subseteq P \text{ for all } j,$$

$$(2.5) \qquad\qquad \bigcup_{j=1}^m P_j \text{ is connected,}$$

$$(2.6) \qquad \begin{array}{l} P_j \text{ and } P_k \text{ for } j \neq k \text{ are either disjoint or intersect at a} \\ \text{single common vertex } S_j \lambda_p = S_k \lambda_q \text{ for some } p, q. \end{array}$$

Note that (2.4) is automatic if the orthogonal transformations R_j are all equal to the identity. Figure 2.1 shows some simple examples: the Sierpinski gasket, the hexagasket and the pentagasket. In contrast, Figure 2.2 shows the octogasket, which is not point–connected because neighboring octagons intersect along edges. In Section 6 we will compute its connectivity dimension. Point–connected sets are included in Kigami's class of post–critically finite (p.c.f.) fractals [K].

For the rest of this section we assume that F is a point–connected self–similar set. We may assume, without loss of generality, that P has non–empty interior, for otherwise we could embed F into a lower dimensional Euclidean space. Note that F satisfies the open set condition with the interior of P as the open set, so the Hausdorff dimension α is determined by the equation

$$(2.7) \qquad\qquad \sum_{j=1}^m \rho_j^\alpha = 1.$$

It is not hard to see that F must be pathwise connected, but this will come out as a consequence of our computation of β.

We begin with an informal description of our construction. Suppose we know the value of β and have a collection of minimal paths γ_{jk} joining λ_j and λ_k, with $\mathcal{H}_\beta(\gamma_{jk}) = L_{jk}$. (In fact it may happen that $L_{jk} = \infty$, but for simplicity we postpone discussion of this case.) We refer to the intersection points in (2.6) as *junction points*. It is clear that any path connecting points in different polytopes
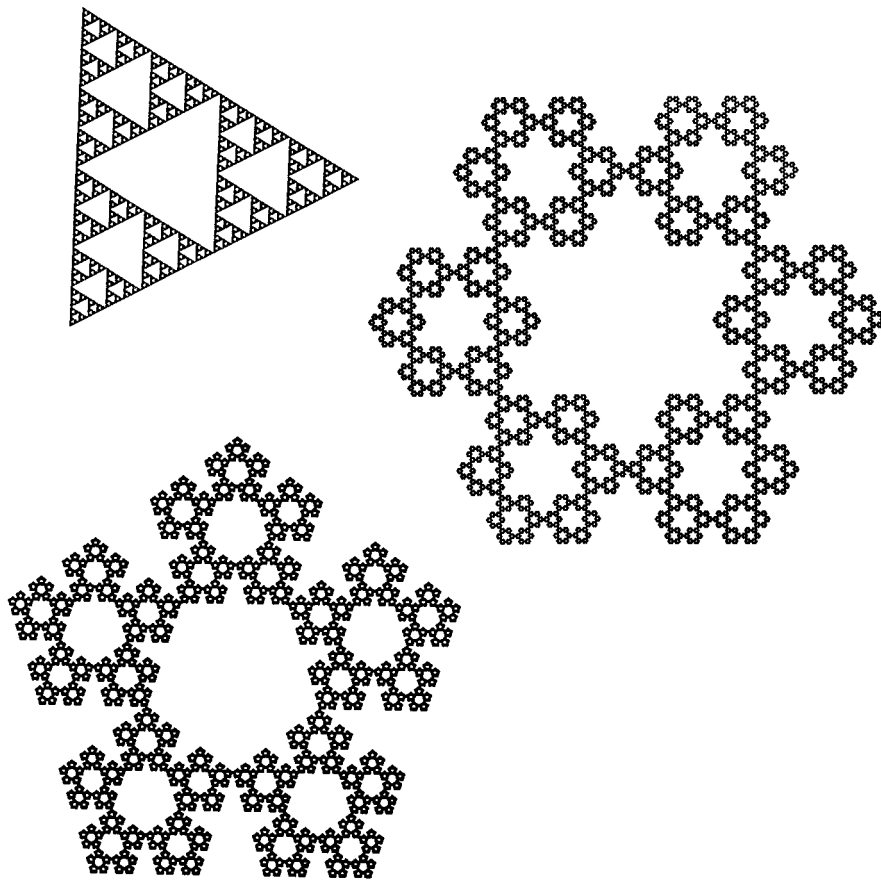
FIGURE 2.1. Some point–connected self–similar sets. (a) The Sier-
pinski gasket. (b) The hexagasket. (c) The pentagasket
.

P_j must pass through junction points to get through, and a minimal path will never pass through a single junction twice. Thus there will be a sequence of $N-1$ distinct junction points $S_{j_\ell}\lambda_{p_\ell} = S_{j_{\ell+1}}\lambda_{q_{\ell+1}}$, $\ell = 1, \dots, N-1$, such that γ_{jk} goes from λ_j to the first junction points, then joins the successive junction points in order, and finally joins the last junction point to λ_k (this means we must have $j_1 = j$ and $j_N = k$). The path γ_{jk} thus breaks up into N pieces, each of which must be a scaled version of one of our original minimal paths (since minimal paths are not necessarily unique, we should not really say "must", but rather "could"). Thus

$$(2.8) \qquad L_{jk} = \rho_j^\beta L_{jp_1} + \sum_{\ell=2}^{N-1} \rho_{j_\ell}^\beta L_{q_\ell p_\ell} + \rho_k^\beta L_{q_N k}$$

by the scaling property of $\mathcal{H}_\beta$. This is the key identity relating the quantities β and L_{jk} that we are trying to compute. We can effectively use it to determine these quantities, and in the process prove that they exist, and furnish a construction of the minimal paths.
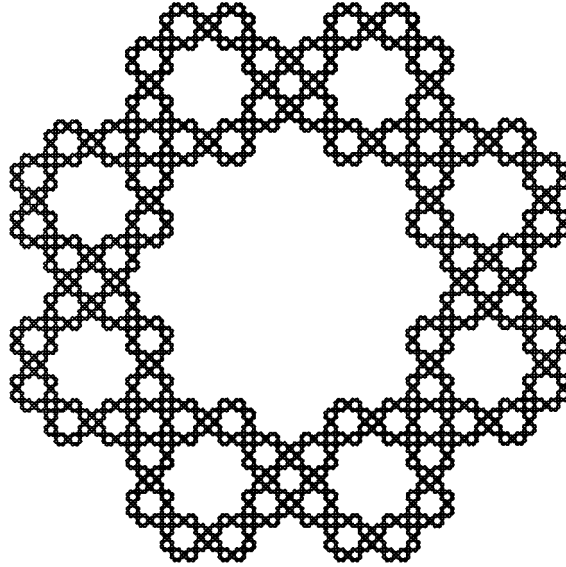
FIGURE 2.2. The octogasket, which is pathwise connected but not a point–connected self–similar set.

But first we have to deal with a problem of ambiguity. There may be many ways, a priori, to connect λ_j with λ_k by adjacent junction points, and we don't know which one (or ones) corresponds to the minimal path. We resolve the ambiguity by doing the computation for all possibilities (there are clearly only a finite number) and choosing the one that gives the smallest values. Let us call a *skeleton path* joining λ_j with λ_k a sequence of distinct adjacent junction points as before. For each choice of skeleton paths for all $j \neq k$, we can write the analogue of (2.8), which yields a matrix equation

$$(2.9) \qquad\qquad M_\beta L = L,$$

where L is the vector with components L_{jk} and the entries of M_β are linear combinations of ρ_i^β with non–negative integer coefficients. Since the entries of the matrix are non–negative and monotone decreasing in β, by the Perron–Frobenius theorem there is a unique value of β for which (2.9) has a unique positive vector solution (unless M_β is reducible, a case we will discuss later), and the value of β is determined by

$$(2.10) \qquad\qquad \text{spec.rad.}(M_\beta) = 1.$$

It is then clear that by iterating the skeleton paths, we obtain in the limit a set of paths γ_{jk} joining λ_j to λ_k with $\mathcal{H}_\beta(\gamma_{jk}) = L_{jk}$. The connectivity dimension is then obtained by taking the minimum value of β over all choices of skeleton paths, and the resulting γ_{jk} are minimal paths. To simplify the calculations we need only consider fixed points that are vertices of the polytope P, since it is only images of these points that can be junction points.

Before stating the main theorem we look at some examples:

Example 2.1. Consider the Sierpinski gasket based on an arbitrary triangle with sides of lengths a, b, c (Figure 2.3 shows the first generation and juncture points).
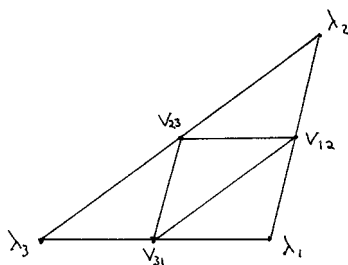
FIGURE 2.3. The first generation of a Sierpinski gasket based on a triangle with side lengths $a = |\lambda_1 - \lambda_2|$, $b = |\lambda_2 - \lambda_3|$ and $c = |\lambda_3 - \lambda_1|$, with the three juncture points labeled.

The obvious skeleton paths connect λ_j with λ_k by going through v_{jk}. This leads to the matrix equation (2.9) for

$$M_\beta = 2^{-\beta} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

In this case (2.10) has the solution $\beta = 1$, showing that the connectivity dimension is at most 1. Since this is the minimum possible, we need look no further. The paths γ_{jk} constructed by this procedure are just the straight lines joining the vertices. Note that we cannot deduce the lengths of these lines from the matrix M_β, because it is not irreducible and has multiple eigenvectors. It is instructive to see what happens if we make a foolish choice of skeleton paths: suppose we make the correct choice for two sides, but join λ_1 to λ_2 by going through v_{12} and then v_{23}. The matrix becomes

$$M_\beta = 2^{-\beta} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and again the solution of (2.10) has $\beta = 1$. The construction produces the straight lines joining λ_1 to λ_2 and λ_2 to λ_3, but the broken line joining λ_1 to λ_3 via λ_2. On the other hand, if we make the foolish choice of skeleton path for all three sides, then the matrix becomes

$$M_\beta = 2^{-\beta} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

and now $\beta = \log 3 / \log 2$ is the solution to (2.10). The construction produces fractal curves joining the vertices. In order to simplify the calculations, we will refrain from considering skeleton paths that are obviously foolish.

Example 2.2. We consider the i.f.s. with 7 contractions with ratio $1/4$ whose first generation is shown in Figure 2.4. The obvious choice for skeleton paths is: $(\lambda_1, v_1, v_2, v_3, \lambda_2)$, $(\lambda_2, v_3, v_4, v_6, v_7, \lambda_3)$, $(\lambda_3, v_7, v_6, v_5, v_1, \lambda_1)$. This yields the
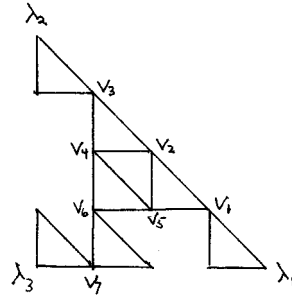
FIGURE 2.4. The first generation in a fractal generated by an i.f.s. with 7 similarities with contraction ratio $1/4$. The three fixed points λ_1, λ_2, λ_3 are the vertices of P. Seven junction points v_j are labeled.
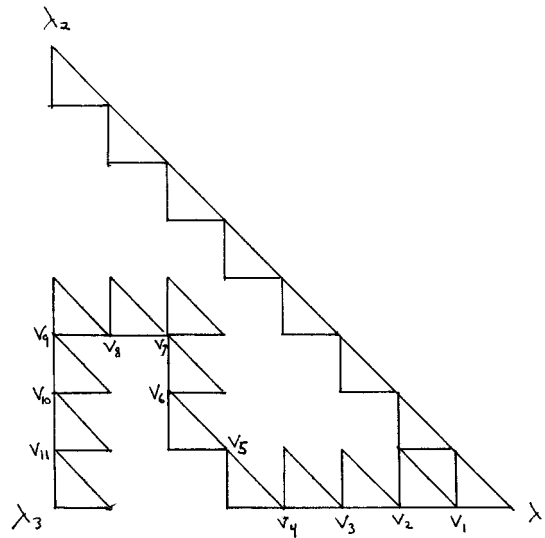

FIGURE 2.5. The first generation for the i.f.s. in Example 2.3, with the three fixed points λ_1, λ_2, λ_3 and some junction points labeled.

matrix

$$M_\beta = 4^{-\beta} \begin{pmatrix} 4 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

and the solution to (2.10) is $\beta = 1$, so the connectivity dimension is 1. However, the construction leads to paths joining λ_2 to λ_3 and λ_3 to λ_1 of infinite (though σ–infinite) length, although λ_1 and λ_2 can be joined by a straight line. In this case M_β is not irreducible, and there is no strictly positive eigenvector corresponding to the eigenvalue 1. In fact there are no paths of finite length connecting λ_2 to λ_3 or λ_3 to λ_1, so we have found the minimal paths.

Example 2.3. We consider the i.f.s. with 20 contractions with ratio $1/8$ whose first generation is shown in Figure 2.5. An obvious choice for skeleton path joins

λ_1 to λ_2 along a straight line, λ_1 to λ_3 passing through $v_1, \dots, v_{11}$ as marked, and λ_2 to λ_3 by the straight line to λ_1 and then the above skeleton path from λ_1 to λ_3. This yields a matrix

$$M_\beta = 8^{-\beta} \begin{pmatrix} 8 & 0 & 0 \\ 2 & 6 & 4 \\ 8 & 4 & 6 \end{pmatrix}$$

and a connectivity dimension of $\log 10 / \log 8$. In this case γ_{13} is a straight line and $\mathcal{H}_\beta(\gamma_{13}) = 0$, but $\mathcal{H}_\beta(\gamma_{13})$ is finite and positive.

In order to state our theorem we need to recall the spectral theory of non–negative matrices and the theory of families of self–similar sets. A square matrix A with non–negative entries is called *irreducible* if some power of A has all entries positive. The Perron–Frobenius theorem says such a matrix has a unique (up to a constant multiple) eigenvector with non–negative entries, and the entries are in fact positive. The corresponding eigenvalue is equal to the spectral radius of A. If A (not identically zero) is not irreducible, then we can transpose the coordinates so that A has the block lower triangular form

$$A = \begin{pmatrix} A_1 & 0 \\ C_1 & D_1 \end{pmatrix}$$

with A_1 irreducible. If it happens that $C_1 = 0$ we can iterate the procedure for D_1. If we end up with A in block diagonal form with each block irreducible, we say that A is *completely reducible*.

Now suppose we are given an oriented graph with vertices V and edges E (each $e \in E$ joins vertices v_j to v_k, and we allow more than one edge to each pair of vertices), and to each $e \in E$ there is a contractive similarity $S(e)$ with contraction ratio $r(e)$. Then the associated family of self–similar sets F_v (one to each vertex) is defined by the system of equations

(2.11) $$F_v = \bigcup_{u \in V} \bigcup_{e \in E_{uv}} S(e)F_u,$$

where E_{uv} denotes the set of edges connecting u to v. There exists a unique solution under the requirement that the sets F_v be compact and nonempty. For such a system the analogue of the open set condition is the existence of open sets U_v such that $S(e)U_u \subseteq U_v$ for $e \in E_{uv}$ and the sets $S(e)U_u$ are disjoint as we vary over all $u \in V$ and all $e \in E_{uv}$, for each fixed v. Associated to this system we form the $|V| \times |V|$ matrix M_β whose (u,v) entry is

(2.12) $$\sum_{e \in E_{uv}} r(e)^\beta.$$

Under these assumptions, the dimension of all the sets F_v is $\leq \beta$, the solution of (2.10). It is easy to see that M_β is irreducible if and only if the oriented graph is connected. If this holds, then $0 < \mathcal{H}_\beta(F_v) < \infty$. For these results see [MW].

Theorem 2.4. *The connectivity dimension of a point connected self–similar set is equal to the minimum value of β solving (2.10) for some set of skeleton paths. If the corresponding matrix M_β is irreducible (for some choice of skeleton paths with minimal β) then the regulairty condition (2.2) holds.*

Proof. Consider the set of fixed points that are vertices of P, say $\lambda_1, \ldots, \lambda_N$. For each distinct pair λ_j, λ_k we choose a path γ_{jk} joining them with dimension $\leq \beta$ and such that $\mathcal{H}_\beta(\gamma_{jk})$ is minimal (perhaps ∞), where β is the connectivity dimension. The existence of such minimal paths follows by a routine compactness argument. Now each path γ_{jk} yields a skeleton path, simply by noting the junction points that γ_{jk} passes through on the way from λ_j to λ_k (in the case that $\mathcal{H}_\beta(\gamma_{jk}) = \infty$ we need a separate argument to show that we can assume that no junction point is repeated). Now we can associate to the set of skeleton paths a family of self–similar sets. The vertices of the directed graph will be the pairs (j, k) with $j < k$, and the set F_{jk} will be paths joining λ_j to λ_k (perhaps not the same as the original γ_{jk}). To each segment of the skeleton path we associate an edge e and a similarity $S(e)$ chosen from the initial set of similarities so that $S(e)$ maps the line segment joining $\lambda_{j'}$ to $\lambda_{k'}$ to the segment of the skeleton path (then e joins the vertex (j', k') to (j, k)). In fact $S(e) = S_\ell$ if the endpoints (a, b) of the segment of the skeleton path lie in $S_\ell P$, and then $S_\ell \lambda_{j'} = a$ and $S_\ell \lambda_{k'} = b$. It is not hard to see that the sets F_{jk} are paths and the dimension of F_{jk} is at most β, with $\mathcal{H}_\beta(F_{jk}) \leq \mathcal{H}_\beta(\gamma_{jk})$. In other words, without loss of generality, we may assume that our paths γ_{jk} form a family of self–similar sets. Furthermore, the open set condition holds with the interior of P as open set. Thus (2.10) holds for this choice of skeleton paths.

This shows that the connectivity dimension is at least the minimum solution to (2.10). But conversely, if β is this minimum solution, then the self–similar family constructed for the corresponding skeleton paths connects all the vertices of P with paths of dimension $\leq \beta$. We claim this implies that any two points in F can be connected by such a path, and for this it suffices to show that λ_1 can be connected to any other point by a path of dimension $\leq \beta$. Let A be the set of all such points. Then we know A contains all the vertices of P, and by iteration it contains all the images of these vertices under iterated mappings from the i.f.s. Thus A is dense in F. But it is easy to see that A is closed, so $A = F$.

Finally, suppose the matrix M_β is irreducible for the choice of skeleton paths leading to minimum paths. Then $0 < \mathcal{H}_\beta(\gamma_{jk}) < \infty$ for these paths, so (2.2) holds if the points x, y are chosen among the vertices of P. It is easy to extend this to the case when just one of the points is a vertex of P. If x and y are arbitrary points on F, we may assume without loss of generality that $x \in S_j P$ and $y \in S_k P$ with $j \neq k$ (for if not we can scale up using inverses of the i.f.s. contractions, since (2.2) is invariant under this scaling). Thus a minimal path γ_{xy} must pass through a junction point v, and by rescaling we return to the case where one of the two points is a vertex of P. It remains to show that $d(x, y)$ is comparable to $d(x, v) + d(v, y)$, and for this we need to use the fact that the image polytopes $S_j P$ intersect only at vertices. This gives a lower bound for the angles of lines from v to the neighboring polytopes, and hence the bound $d(x, v) + d(v, y) \leq cd(x, y)$.   Q.E.D

There are other cases when the regularity estimate (2.2) holds when the matrix M_β is not irreducible, as Example 2.1 shows, but we will not attempt to state a formal theorem to this effect. In our applications there will be a number of symmetries that imply isometries between some of the minimal paths, and thus enable us to reduce the order of the matrices and simplify the computations.

When $\beta = 1$, we can compute the exact length of the minimal paths. In this case the paths are the limits of a sequence of broken line segments, and the length of a limiting path is the limit of the lengths of the broken line segments. We start with

straight lines connecting the vertices, so we have an initial vector $L_{jk} = d(\lambda_j, \lambda_k)$ of lengths, and at each iteration of the construction we multiply this vector by M_1. Thus $\lim\limits_{N \to \infty} M_1^N L$ gives the lengths of the minimal paths. Some of these may be infinite, as in Example 2.2. In general, when $\beta > 1$ and the matrix M_β corresponding to the minimal paths is irreducible, we can only determine the $\mathcal{H}_\beta$ measures of the minimal paths up to a constant multiple by the observation that (2.9) holds for K equal to the vector with $L_{jk} = \mathcal{H}_\beta(\gamma_{jk})$, and the Perron–Frobenius theorem. In general it is quite difficult to compute exact Hausdorff measures (see [AS], [Ma1], [Ma2] for some special cases where it is possible).

We now carry out the computation of connectivity dimension for the general class of regular polygaskets. We take for P a regular polygon in $\mathbb{R}^2$ with N sides (specifically, we take the vertices $\lambda_j = \omega^j$ where ω is a primitive N–th root of unity, and $\mathbb{R}^2$ is identified with $\mathbb{C}$). For each vertex ω^j we take S_j with fixed point ω^j and contraction ratio

$$(2.13) \qquad r = \frac{\sin \pi/N}{\sin \pi/N + \sin(\pi/N + 2\pi m/N)}$$

for $m = [N/4]$. This choice of r makes the polygons $S_j P$ just touch each other, at a single vertex if N is not divisible by 4, and along a single edge if N is divisible by 4. We call the resulting fractal the N–polygasket, denoted PG_N. Some of these ($N = 3, 5, 6, 8$) are illustrated in Figures 2.1 and 2.2. Note that PG_4 is just a square. When N is not divisible by 4 we have a point–connected self–similar set.

Theorem 2.5. *The connectivity dimension β of PG_N for N not divisible by 4 is given by*

$$(2.14) \qquad r^{-\beta} = \begin{cases} m + \sqrt{m^2 - 2m + 4}, & N = 4m + 1, \\ m + \dfrac{1}{2} + \sqrt{m^2 - m + 9/4}, & N = 4m + 2, \\ m + 1 + \sqrt{m^2 + 1}, & N = 4m + 3. \end{cases}$$

The regularity estimate (2.2) holds. The only cases with $\beta = 1$ are $N = 3$ or 6.

Proof. By symmetry, all minimal paths γ_{jk} joining ω^j and ω^k are isometric if $j - k$ is equal mod N. However, we can reduce the order of the matrix M_β to 2×2 by considering just two such paths, A joining ω^0 and ω^{m+1}, and B joining ω^0 and ω^{N-2m-2}, and their measures $a = \mathcal{H}_\beta(A)$, $b = \mathcal{H}_\beta(B)$. (In fact, for $N = 3$, 6 or 9, A and B are identical, but this does not affect the rest of the argument.) The key geometric observation is that every minimal path connecting adjacent junction points is the image under S_j of a path isometric to B, while every minimal path connecting a vertex ω^j to one of the two junction points in $S_j P$ is the image under S_j of a path isometric to A. This is apparent in the cases illustrated in Figure 2.1, and can be proved in general by noting that the junction points in $S_0 P$ are $S_0 \omega^{m+1}$ and $S_0 \omega^{N-m-1}$, since $S_0 \omega^{m+1} = S_1 \omega^{-m}$ and $S_0 \omega^{N-m-1} = S_{N-1} \omega^m$. Thus $S_0 A$ connects $\omega^0 = S_0 \omega^0$ to the junction point $S_0 \omega^{m+1}$, while $S_0 \widetilde{B}$ ($\widetilde{B}$ a rotation of B) connects the junction points $S_0 \omega^{m+1}$ and $S_0 \omega^{N-m-1}$.

Now the obvious skeleton paths corresponding to A and B proceed counterclockwise (the clockwise alternative is longer), joining the initial vertex ω^0 to the first junction point $S_0 \omega^{m+1}$, joining adjacent junction points (m or $N - 2m - 3$ in number) and finally joining the last junction point to the destination vertex (ω^{m+1} or ω^{N-2m-2}). Thus A consists of 2 scaled images of A and m scaled images of B,

while B consists of 2 scaled images of A and $N - 2m - 3$ scaled images of B. Thus

$$(2.15) \qquad M_\beta = r^\beta \begin{pmatrix} 2 & m \\ 2 & N - 2m - 3 \end{pmatrix}.$$

The eigenvalues of M_β are just r^β times the roots of the quadratic

$$(\lambda - 2)(\lambda - (N - 2m - 3)) - 2m = 0,$$

and choosing the larger root we obtain the solution of (2.10) in the form given by (2.14). Since M_β is irreducible we obtain the regularity estimate (2.2). From (2.9) we find the relative size of a and b, namely

$$(2.16) \qquad b/a = (1 - 2r^\beta)/mr^\beta.$$

It is easy to find the relative sizes of the $\mathcal{H}_\beta$ measures of the minimal paths connecting any two vertices. For $k \le N/2$, a minimal path connecting ω^0 and ω^k corresponds to a skeleton path with 2 scaled copies of A and $k - 1$ scaled copies of B; hence the $\mathcal{H}_\beta$ measure is

$$(2.17) \qquad r^\beta(2a + (k-1)b) = \left(2r^\beta + \frac{k-1}{m}(1 - 2r^\beta)\right)a.$$

For large values of N, we have

$$\frac{1}{r} = 1 + \frac{\sin \frac{2m+1}{N}\pi}{\sin \frac{\pi}{N}} \approx 1 + \frac{N}{\pi},$$

while (2.14) says $r^{-\beta} \approx N/2$, hence

$$(2.18) \qquad \beta \approx 1 + \frac{\log \pi/2}{\log N} \quad \text{as } N \to \infty.$$

Thus $\beta = 1$ is possible for only small N, and a direct check reveals that this happens only for $N = 3$ and 6. $\qquad\qquad$ Q.E.D

When N is divisible by 4, we cannot use Theorem 2.4 to compute the connectivity dimension. Nevertheless, if we were to follow the same pattern, we would have

$$(2.19) \qquad r^{-\beta} = m - 1/2 + \sqrt{m^2 - 3m + 25/4}$$

for $N = 4m$ ($m \ge 2$). In Section 6 we will verify that this is correct when $N = 8$, but the proof is much more intricate.

3. Isoperimetric estimates for the Sierpinski gasket

The connectivity dimension of the Sierpinski (SG) gasket is one. We define the *isoperimetric profile function* $h(L)$ to be the supremum of the measure of a region bounded by a curve of length at most L. Thus

$$(3.1) \qquad V(\Omega) \le h(L) \text{ if } \Omega \text{ is bounded by } \gamma \text{ with } L(\gamma) \le L.$$

It is clear by scaling that

$$(3.2) \qquad h(L) \le cL^\alpha \quad \text{for} \quad \alpha = \log 3/\log 2,$$

for some constant c. Our goal is to find the best constant. In fact we will see that it is $24/(21)^\alpha$, and equality holds for the region shown in Figure 3.1, and only for regions that are similar.

The main idea we use is to consider, for each closed curve γ, the largest deleted triangle it contains. Suppose first that this is the big deleted triangle of side length $1/2$. We call such curves *big curves*. Then the length of γ is at least $3/2$. But
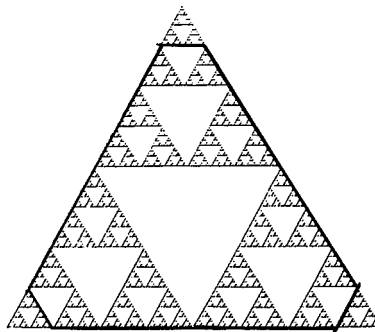
FIGURE 3.1. The curve of length 21/8 bounding a region of measure 8/9. This region yields the best constant in (3.2), and up to similarities it is the unique such region.
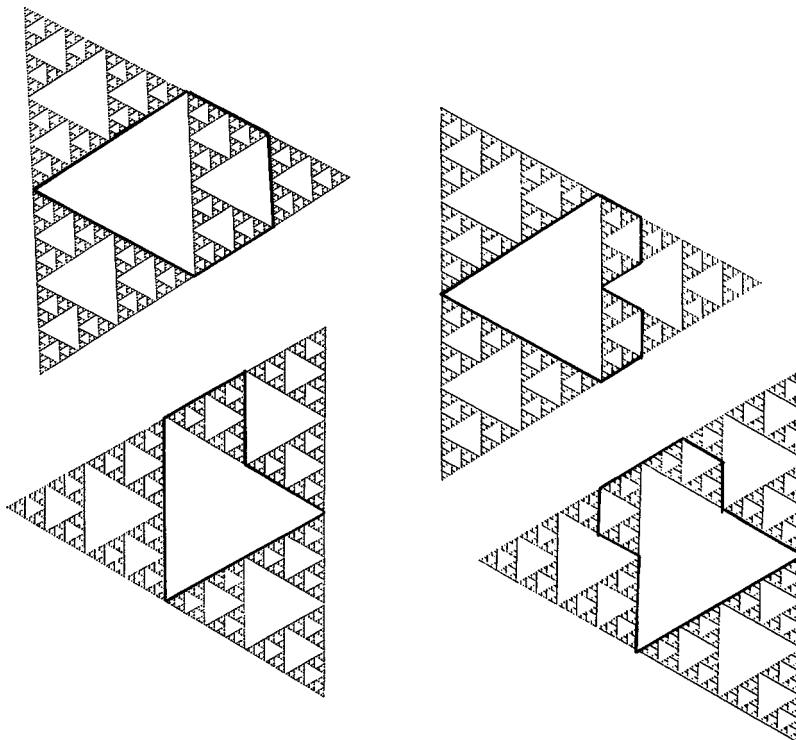

FIGURE 3.2. Four curves of length 7/4. Clearly the first one bounds the region of greatest measure, 2/9.

there is already an obvious choice of a curve, namely one of the triangles in SG of side length $1/2$, whose length is $3/2$, bounding a region of measure $1/3$. So for computing $h(L)$, we need only consider those curves bounding regions of measure greater than $1/3$ (among big curves). We will soon see how to raise this value to $4/9$.
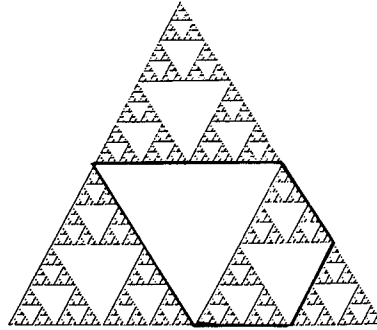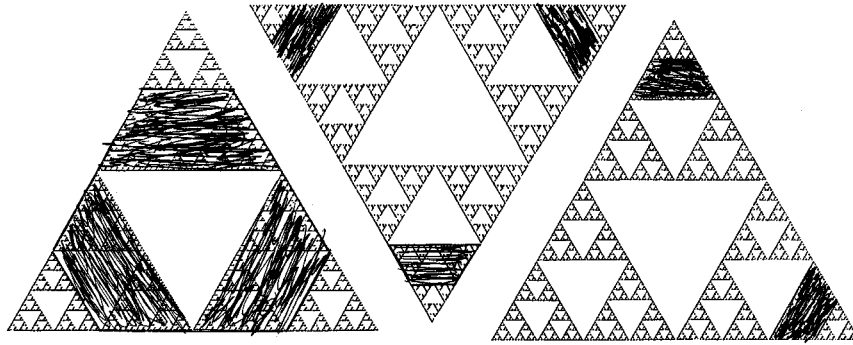
FIGURE 3.3. The smallest big region that we need to consider.

FIGURE 3.4. The hatched areas are the shells. a) The shell of order 1. b) The shell of order 2. c) A 2/3 partial shell of order 2.

The next idea is illustrated in Figure 3.2. All four curves shown have the same length, 7/4, but the first one bounds the region of greatest measure, 2/9. The moral of the story is that it doesn't make sense to use curves that bite off a single small triangle of a pair (as the second one does), or two adjacent 2/3 triangles (as the third one does). The fourth curve shows that the two 2/3 triangles need not be adjacent. The same is true on smaller scales as well.

Now let us return to considering a big curve γ, and the region Ω it bounds. At least one of the 3 first generation triangles must contain a portion of Ω of measure at least 1/9 (since the total measure is at least 1/3), and by the above considerations the minimum size must be as shown in Figure 3.2 (a). Of course this curve is not reasonable, because it has length 7/4 but only bounds a region of measure 2/9. We need to add a measure of at least 1/9, and by the above it will cost us no more in length to add a measure of 2/9, as in Figure 3.3. This curve, of length 2, bounds a region of measure 4/9, and is the smallest big curve that we need consider. Thus, we will be looking at big curves whose lengths satisfy $2 \leq L \leq 3$, bounding regions of measure $4/9 \leq V \leq 1$. It is conceivable that there might be curves in this length range that are not big but still bound a region of greater measure, but we will eventually show that this does not happen.

Now we can describe exactly the curves of lengths satisfying $2 \leq L \leq 3$ that bound the regions of maximum measure. We think of the region as being built up
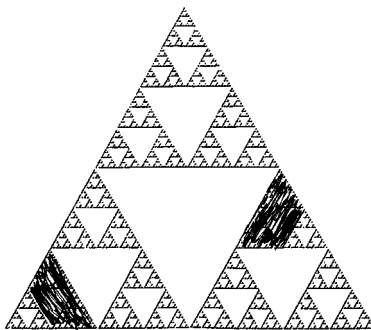
FIGURE 3.5. Two different positions for a mound of order 2.

from the big deleted triangle by adding on certain *shells*. The full shells of orders 1 and 2 are illustrated in Figure 3.4. Each shell consists of 3 pieces, one in each of the three first generation triangles, and we will need to consider a 2/3 partial shell consisting of just 2 of these pieces. We also need to consider just one of these pieces, which we call a *mound*, but unlike the case of a shell we will allow the mound to occupy many different positions, as shown in Figure 3.5. It is clear that adding a shell of order k increases the measure by $2 \cdot 3^{-k}$ and the length by $3 \cdot 2^{-k+1}$, and these values are reduced by factors of 2/3 and 1/3 for a 2/3 partial shell and a mound, respectively.

A type I region is determined by specifying an integer $m \geq 1$ and a subset B of the integers $> m$, whose complement is infinite, and consists of the shells of orders $1, 2, \ldots, m$ together with one mound of order k for each $k \in B$. (There is quite a bit of choice as to the placement of the mounds, but this does not affect either the measure of the region or the length of the bounding curve.) A type II region is determined by the same data, and consists of the shells of orders $1, 2, \ldots, m-1$ (none if $m = 1$), a 2/3 partial shell of order m, and one mound of order k for each $k \in B$.

In order to describe the measures of these regions and the lengths of their bounding curves, it is convenient to associate with the set B the non–negative number $x < 2^{-m}$ whose base 2 expansion has a 1 in the k–th place for each $k \in B$. (The requirement that the complement of B be infinite means that the expansion of x does not end with an infinite string of 1's, and this resolves the ambiguity in the expansion of binary rational numbers.) We define a function $\psi(x)$ to be the same expansion interpreted as a base 3 number. This means

$$(3.3) \qquad \psi\Big(\sum_{k \in B} 2^{-k}\Big) = \sum_{k \in B} 3^{-k}.$$

Given the data m and B (hence the number $x < 2^{-m}$), we find that the measure V and length L for a type I region are given by

$$(3.4) \qquad 1 - 3^{-m} \leq V = 1 - 3^{-m} + (2/3)\psi(x) < 1 - 2 \cdot 3^{-m-1},$$

$$(3.5) \qquad 3 - 3 \cdot 2^{-m-1} \leq L = 3 - 3 \cdot 2^{-m-1} + (1/2)x < 3 - 2^{-m},$$

and for a type II region by

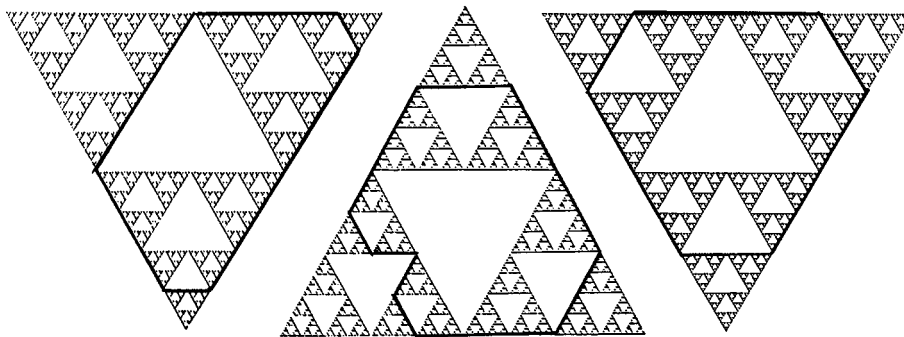$$(3.6) \qquad 1 - 5 \cdot 3^{-m-1} \leq V = 1 - 5 \cdot 3^{-m-1} + (2/3)\psi(x) < 1 - 4 \cdot 3^{-m-1},$$

FIGURE 3.6. (a) shows a curve with two 2/3 partial shells, of orders 1 and 2. (b) shows an equivalent region, with the same $L(9/4)$ and $V(16/27)$. (c) shows a region with the same L and larger $V(2/3)$.

$$(3.7) \qquad 3 - 2^{-m-1} \le L = 3 - 2^{-m+1} + (1/2)x < 3 - 3 \cdot 2^{-m-1}.$$

Note that the intervals for L in (3.5) and (3.7) are non–overlapping and fill the range $2 \le L < 3$. (The case $L = 3$ corresponds clearly to the whole triangle with $V = 1$, which we can include as the limiting case as $m \to \infty$.)

Theorem 3.1. *For $2 \le L < 3$, the type I or II region bounded by a curve of length L maximizes the measure over all regions bounded by curves of length $\le L$. Thus $h(L)$ is defined by (3.4) and (3.5), or by (3.6) and (3.7), depending on the value of L. In particular, $h(L)$ is a discontinuous function.*

Theorem 3.2. *The function $h(L)$ satisfies (3.2) for all $L \le 3$ with $c = 24/(21)^\alpha \approx .1925498$, with equality holding for $L = 21/8$ and $V = 8/9$ for the region shown in Figure 3.1, a type I region with $m = 2$ and B empty. Furthermore, equality holds only for similar regions.*

The proofs of the two theorems are intertwined. We have already proved most of Theorem 3.1 for regions bounded by big curves. The configurations illustrated in Figure 3.2 eliminate from consideration all big curves except those of type I or II. Indeed, (c) and (d) illustrate why we can never add two or more mounds of the same order, unless they are part of a complete shell or a 2/3 partial shell. The folly of taking two 2/3 partial shells of different order is illustrated in Figure 3.6. Similar reasoning shows we must take all complete shells in consecutive order, and if we take a single 2/3 partial shell it must immediately follow a complete shell. Finally, we need to explain the condition that the complement of B be infinite. This means we rule out adding on a complete triangle. In Figure 3.7 we illustrate what this would entail, and why it can be ruled out. Note that the case illustrated in Figure 3.7 (b) corresponds to taking $x = 2^{-m}$ in (3.5) or (3.7), and this always loses in competition with taking $x = 0$ in the other (for the appropriate m value), since we obtain the same L value with larger V. Here we see clearly the discontinuity of the function $h(L)$. In fact, the set of discontinuities is dense because the function $\psi(x)$ is discontinuous at all dyadic rationals.

This completes the proof of Theorem 3.1 if we can show that only big regions contribute to $h(L)$ for $2 \le L < 3$. Before we can show this we need to begin work
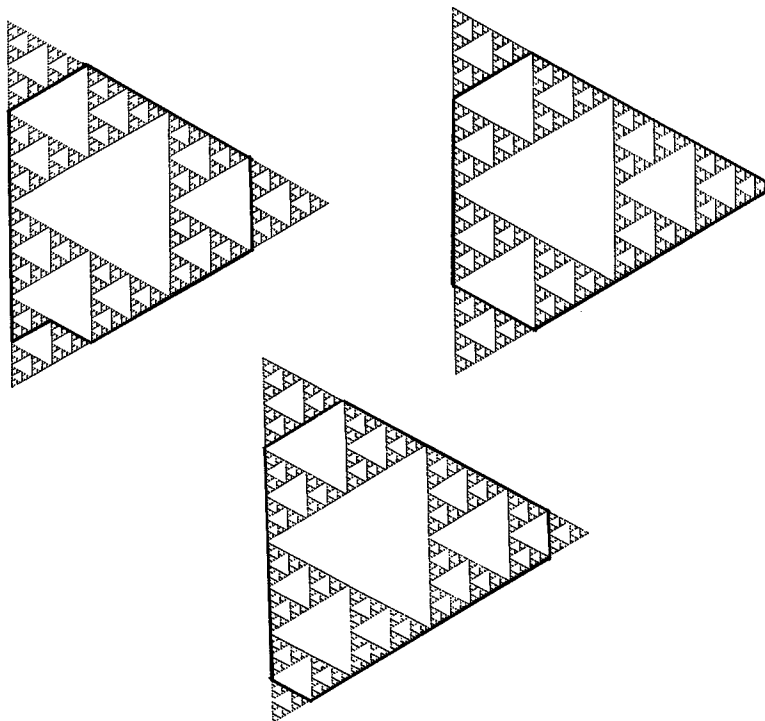
FIGURE 3.7. (a) A type I region with $m = 1$ and $B = \{3, 4, \ldots\}$, which adds a small triangle to a shell. We know from Figure 3.2 (b) why this is foolish. (b) A type I region with $m = 1$ and $B = \{2, 3, \ldots\}$. This has the same L but smaller V than the type II region with $m = 1$ and B empty shown in (c).

on the proof of Theorem 3.2. The following lemma is an immediate consequence of (3.3).

Lemma 3.3. *Let* $\alpha = \log 3 / \log 2$. *Then*

$$(3.8) \qquad\qquad \psi(x) \leq x^\alpha \text{ for } 0 \leq x \leq 1.$$

We use the estimate (3.8) in (3.4) and (3.6), to obtain

$$(3.9) \qquad\qquad V \leq 1 - 3^{-m} + 2(x/2)^\alpha$$

and

$$(3.10) \qquad\qquad V \leq 1 - 5 \cdot 3^{-m-1} + 2(x/2)^\alpha$$

in each case, and we note that equality holds at the endpoints $x = 0$ and $x = 2^{-m}$. We claim that the function V/L^α on each interval defined by (3.5) or (3.7) (for one fixed m) attains its maximum on one of the endpoints, and to do this it suffices to prove the same when we replace V by the upper bound given in (3.9) and (3.10). Since the logarithm is increasing, this is equivalent to showing that

$$\log(1 - 3^{-m} + 2(x/2)^\alpha) - \alpha \log(3 - 3 \cdot 2^{-m+1} + (1/2)x)$$

and

$$\log(1 - 5 \cdot 3^{-m-1} + 2(x/2)^\alpha) - \alpha \log(3 - 2^{-m+1} + (1/2)x)$$

are concave up on $0 \le x \le 2^{-m}$, which is a simple exercise in calculus (in fact they are concave up on $0 \le x$, and the key fact is that $\alpha > 1$).

Thus, to find the maximum of V/L^α we need only examine the values of the end-points of the intervals in (3.5) and (3.7), and we have already observed that the upper endpoints will not do. Thus the problem reduces to maximizing $(1 - 3^{-m})/(3 - 3 \cdot 2^{-m-1})^\alpha$ and $(1 - 5 \cdot 3^{-m-1})/(3 - 2^{-m+1})^\alpha$. We can make this into a calculus problem by setting $x = 2^{-m}$ and $x^\alpha = 3^{-m}$, and we find the maximum for the first function at $x = \left(\frac{1}{2}\right)^{1/(\alpha-1)}$, and for the second function at $x = \left(\frac{2}{5}\right)^{1/(\alpha-1)}$. The final contest is between $m = 1$ or 2 in the first and $m = 2$ or 3 in the second, and $m = 2$ in the first wins.

Again this does not quite complete the proof of Theorem 3.2, since we have only found the maximum for V/L^α in the range $2 \le L < 3$ and for big regions. But now we can use this to complete the proof of Theorem 3.1. Suppose we had a region that was not big with $L \ge 2$ and $V \ge 4/9$. Then it would be the union of 2 or 3 regions, each contained in a single first generation triangle. Suppose it were the union of 2 regions, each of which is big when scaled up to the whole gasket. Then we would have $V = V_1 + V_2$ and $L = L_1 + L_2$ with $V_1 \le cL_1^\alpha$ and $V_2 \le cL_2^\alpha$ by the portion of Theorem 3.2 already established, since (3.2) scales appropriately.

We consider first the case $L = 2$. The maximum value of $L_1^\alpha + L_2^\alpha$ is then $\left(\frac{3}{2}\right)^\alpha + \left(\frac{1}{2}\right)^\alpha$, since it would not make sense to take $L_1 > 3/2$. We then have $V \le c\left(\left(\frac{3}{2}\right)^\alpha + \left(\frac{1}{2}\right)^\alpha\right) = \frac{8(3^\alpha+1)}{(21)^\alpha} \approx .4303$, which is less than $4/9$, the value for V corresponding to $L = 2$ for big regions. In fact, if we take $L = 2 + x/2$, the same estimate shows $V \le 4/9$ provided $x \le .1336$, so the big regions win in this range. But for $x \ge .1336$, we have $\psi(x) \ge 1/27$, so the big regions can produce a measure of at lest $38/81$, and we can take x up to $.3479$ and have our upper bound $c\left(\left(\frac{3}{2}\right)^\alpha + \left(\frac{1+x}{2}\right)^\alpha\right)$ come short of $38/81$, so the big regions still win. Finally, for $x \ge .3479$ we have $\psi(x) \ge 1/9$, so the big regions can produce a measure of at least $14/27$, and this beats our upper bound for x up to $1/2$, which is as far as we have to go, since for $L = 9/4$ the big region has measure $2/3$, which is the total measure of two complete first generation triangles.

Similar reasoning shows that we cannot win with 3 or more small regions. This justifies our limitation to big regions for $2 \le L < 3$, and completes the proof of Theorem 3.1.

Now we can complete the proof of Theorem 3.2. We already have the estimate for big regions, and by scale invariance for any region similar to a big region. It is easy to see that any region is a finite or countable union of regions similar to a big region. Since $\alpha > 1$ we see that $V_j \le cL_j^\alpha$ implies $\sum V_j \le c \sum L_j^\alpha \le c(\sum L_j)^\alpha$, so the estimate holds for all regions. Also, we cannot have equality unless there is only one non–zero sumand; in other words, the region is similar to a big region. But we have already seen that for big regions there is only one that achieves equality.

Theorem 3.1 allows us to compute the isoperimetric profile function for values of L in the range $2^{-k+1} \le L \le 3 \cdot 2^{-k}$ for $k = 0, 1, 2, \ldots$ by scaling, but not for other values of L. In particular, for $3/2 < L < 2$ the situation seems to be quite complicated.
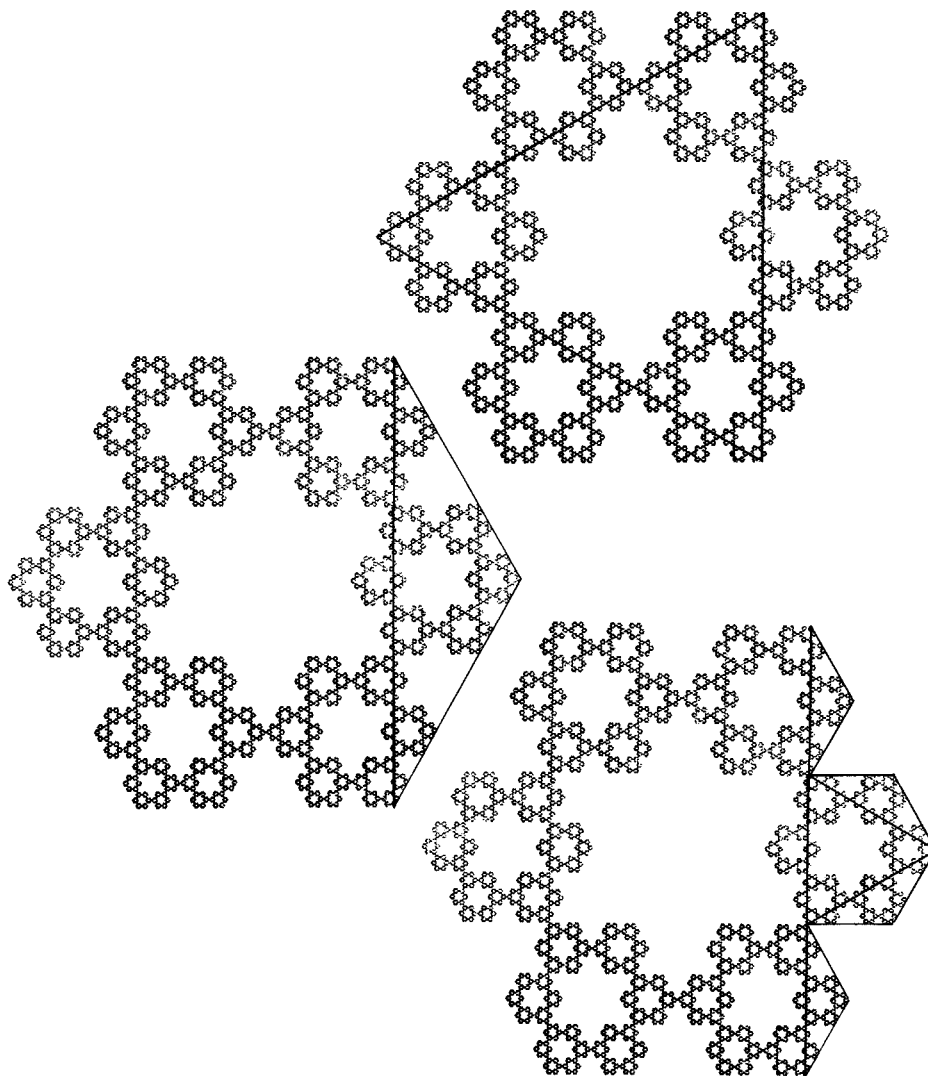
FIGURE 4.1. (a) The triangle T. (b) The wedge W. (c) W dis-
sected into 4 wedges and 1 triangle of order 1.

4. ISOPERIMETRIC ESTIMATES ON THE HEXAGASKET

The hexagasket (PG_6) has connectivity dimension 1, but unlike the Sierpinski
gasket it has no bounding curve of finite length. This means the isoperimetric profile
function is of interest for all positive values of L, and we will study its asymptotic
behavior as $L \to \infty$, which does not scale in the same way as it does for small
values of L.

Our first task is to determine the measures of the basic triangles that will be the
building blocks of our regions: the triangle T and the wedge W shown in Figure 4.1.
More generally, we call a *triangle of order k* any image of T under k similarities,
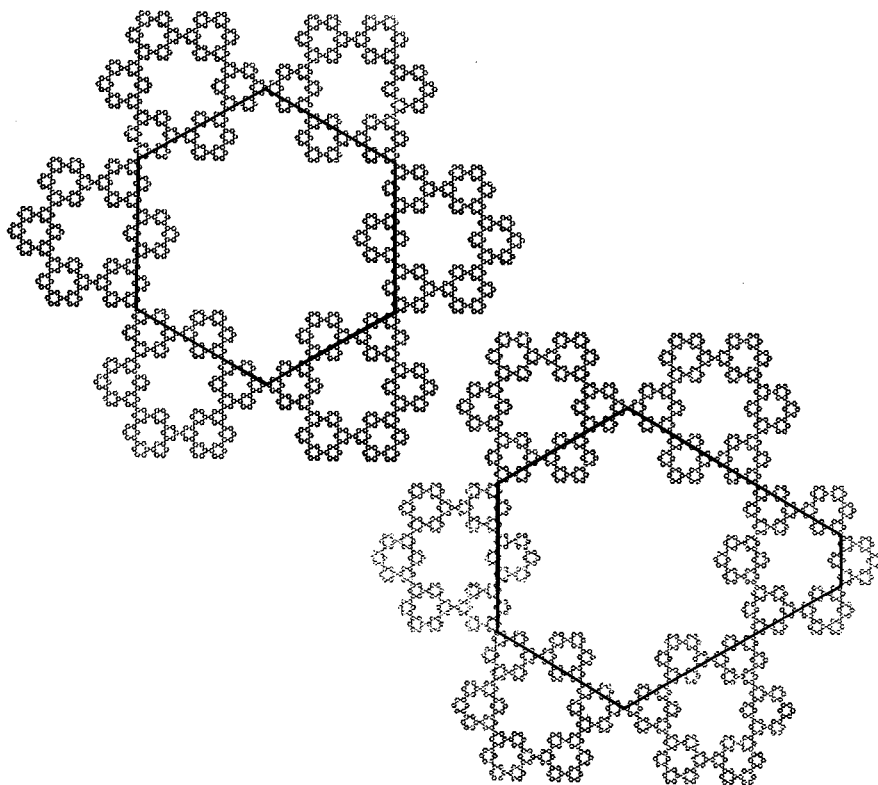and a *wedge of order k* any image of W under k similarities.

FIGURE 4.2. (a) The hexagon, the big region with shortest perimeter. It contains 6 wedges of order 1. (b) The smallest region containing the hexagon and 1 deleted snowflake of order 1. The added volume amounts to 1 hexagon of order 1, and 2 triangles of order 2.

Lemma 4.1. *The measure of W is 1/5, and the measure of T is 2/5.*

Proof. Since T is the complement of three wedges, it suffices to find the measure of W. However, it is clear from Figure 4.1 (c) that W is the union of 4 wedges of order 1 and 1 triangle of order 1. Thus the measure w of W satisfies

$$w = \frac{4}{6}w + \frac{1}{6}(1 - 3w);$$

hence $w = 1/5$.   Q.E.D

Each side of T has length $\sqrt{3}$, and each side of a triangle of order k has L equal to $3^{-k}\sqrt{3}$.

We call a region *big* if it contains the big deleted snowflake in the center of the hexagasket. The shortest curve bounding a big region is clearly the hexagon shown in Figure 4.2. It has $L = 2\sqrt{3}$ and $V = 1/5$ (6 wedges of order 1). We also show in Figure 4.2 the smallest big region obtained by adjoining this hexagon to a single deleted snowflake of order 1. This process increases L by $(2/9)\sqrt{3}$ and V by 1/18 (1 hexagon of order 1 and 2 triangles of order 2). We could use this process to add a variable number of deleted snowflakes of order 1 (up to 6), and then variable

numbers of deleted snowflakes of higher order. It is clear that these are the only big regions we need to consider in computing $h(L)$, although not all of them will be optimal. If we let a_k denote the number of deleted snowflakes of order k that we adjoin, we have

$$(4.1) \qquad L = \left(2 + 2\sum_{k=1}^{\infty} a_k 3^{-k-1}\right)\sqrt{3}$$

and

$$(4.2) \qquad V = \frac{1}{5} + 2\sum_{k=1}^{\infty} a_k 6^{-k-1}$$

for these regions. It is important to know exactly which values of a_k are allowed.

Lemma 4.2. *Any sequence* $a_1, a_2, \ldots$ *satisfying*

$$(4.3) \qquad 0 \le a_k \le 3^k + 3 + \sum_{j=1}^{k-1} a_j 3^{k-1-j}$$

is allowed. In particular, the largest sequence has

$$(4.4) \qquad a_k = 4^k + 2.$$

Proof. Consider the sequence of regions obtained by adjoining to the hexagon the a_j deleted snowflakes of order j for $j < k$. These regions are bounded by polygons made up of line segments of length $\sqrt{3}\,3^{1-k}$, and these line segments provide the opportunity to continue the construction to the next stage. We note from Figure 4.2 that there are two types of segments, the *innies* and the *outies*, depending on whether the adjacent deleted snowflake of order k lies within or outside the region. Let b_k and c_k denote the number of innies and outies at each level. It is clear that a_k may take any value between 0 and c_k, so to prove (4.3) we need to show that the right side is equal to c_k. To do this we need to find the recursion relations for these quantities.

Initially we have $b_1 = 0$, $c_1 = 6$. At stage k in the construction, each line segment either comes from subdividing into 3 one of the segments from the previous stage, or from the replacement of one of the segments of the previous stage by 5 segments as illustrated in the passage from Figure 4.2 (a) to 4.2 (b). We see from the figure that each innie subdivides into 2 innies and 1 outie, each outie subdivides into 1 innie and 2 outies, and the replacement of each outie yields 2 innies and 3 outies. There are b_{k-1} innies that all subdivide, and of the c_{k-1} outies, a_{k-1} get replaced and $c_{k-1} - a_{k-1}$ subdivide. This leads to the recursion relations

$$(4.5) \qquad \begin{cases} b_k = 2b_{k-1} + (c_{k-1} - a_{k-1}) + 2a_{k-1} = 2b_{k-1} + c_{k-1} + a_{k-1}, \\ c_k = b_{k-1} + 2(c_{k-1} - a_{k-1}) + 3a_{k-1} = b_{k-1} + 2c_{k-1} + a_{k-1}. \end{cases}$$

These are easily solved by adding and subtracting. We find that $c_k - b_k$ is constant; hence $c_k - b_k = 6$, and $c_k + b_k = 3(c_{k-1} + b_{k-1}) + 2a_{k-1}$; hence

$$c_k + b_k = 2\left(3^k + \sum_{j=1}^{k-1} a_j 3^{k-1-j}\right),$$

which yields the right side of (4.3) for c_k. Finally, we get the largest value for a_k by taking the largest value for all a_j with $j < k$; hence $a_k = c_k$ and equality holds on the right side of (4.3), and then (4.4) follows by induction.   Q.E.D

We note that the choice of (4.4) for a_k yields $V = 1$ in (4.2), and this is the only choice for which $V = 1$. We, of course, get ∞ for L in (4.1) for this choice, as well as many other choices. For each fixed $L \geq 2\sqrt{3}$, we want to find the choice of a_k that makes (4.1) hold and maximizes V in (4.2). The key observation is that we want to make the values of a_k as large as possible for smaller values of k, since if we reduce a_{k+1} by 3 and increase a_k by 1 we will leave L unchanged and increase V by 6^{-k-1}.

By repeating this process (infinitely often, if need be) we arrive at the following type of region: there is a positive integer m such that $a_k = 4^k + 2$ for $k < m$,

(4.6) $$0 \leq a_m < 4^m + 2,$$

$a_k = 0$, 1 or 2 for $k > m$, but we exclude $a_k = 2$ for all $k \geq N$ for every N. The data for such a region is m, a_m satisfying (4.6) and a real number x satisfying $0 \leq x < 3^{-m}$ with base 3 representation

(4.7) $$x = \sum_{k=m+1}^{\infty} a_k 3^{-k}$$

(with the understanding that ambiguities are resolved in favor of infinite strings of 0's over infinite strings of 2's). In this case (4.1) and (4.2) become

(4.8)
$$\frac{6}{\sqrt{3}}\left(\frac{4^m - 1}{3^{m-1}} + \frac{a_m}{3^m}\right) \leq L = \frac{6}{\sqrt{3}}\left(\frac{4^m - 1}{3^{m-1}} + \frac{a_m}{3^m} + x\right) < \frac{6}{\sqrt{3}}\left(\frac{4^m - 1}{3^{m-1}} + \frac{a_m + 1}{3^m}\right)$$

and

(4.9)
$$1 - \frac{4}{5 \cdot 6^m} - \left(\frac{2}{3}\right)^m + \frac{2a_m}{6^{m+1}} \leq V$$
$$= 1 - \frac{4}{5 \cdot 6^m} - \left(\frac{2}{3}\right)^m + \frac{2a_m}{6^{m+1}} + \frac{\psi(x)}{3}$$
$$< 1 - \frac{4}{5 \cdot 6^m} - \left(\frac{2}{3}\right)^m + \frac{2a_m + 1}{6^{m+1}},$$

where $\psi(x)$ interprets x as a base 6 number, so

(4.10) $$\psi\left(\sum a_k 3^{-k}\right) = \sum a_k 6^{-k}.$$

Note that the intervals in (4.8) exactly fill the range $L \geq 2\sqrt{3}$, and so we can use (4.8) and (4.9) to define a function $h(L)$. This function is evidently discontinuous.

In order to show that this is indeed the isoperimetric profile function, we need to rule out regions that are not big. We will not present all the details, because it is not a plausible possibility that unions of scaled down images of big regions could compete with big regions.

Now let $\alpha = 1 + (\log 2 / \log 3)$, the dimension of the hexagasket. We want to determine the best constant in the inequality

(4.11) $$V \leq cL^\alpha.$$

Theorem 4.3. *The best constant in (4.11) is*

(4.12) $$c = \frac{43}{90} \cdot \left(\frac{\sqrt{3}}{28}\right)^\alpha \approx .0051060895,$$

and this value is attained for the region shown in Figure 4.3, corresponding to data $m = 1$, $a_1 = 5$, $x = 0$, and only for regions similar to it.
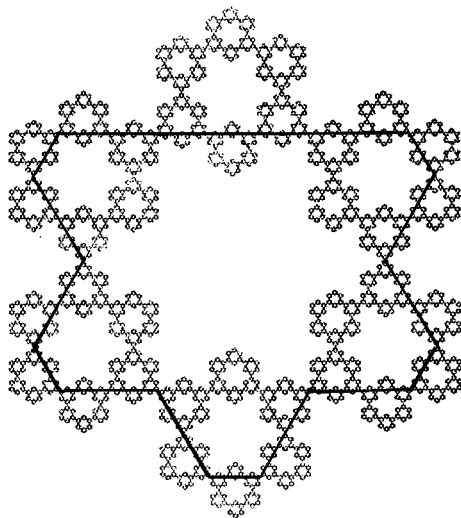
FIGURE 4.3. The extremal region for the isoperimetric inequality
(4.11), with $L = 28/\sqrt{3}$ and $V = 43/90$.

Proof. It suffices to prove this for big regions since (4.11) scales appropriately under
similarities; and since $\alpha > 1$ it holds for unions of regions once we have proved it
for the summands.

The problem is to maximize V/L^α, or equivalently $\log V - \alpha \log L$, where L and
V satisfy (4.8) and (4.9). Our first observation is that it suffices to consider $x = 0$.
Indeed from (4.10) we have

(4.13) $\psi(x) \leq x^\alpha,$

with equality holding if $x = 0$ or $x = 3^{-m}$. Thus it suffices to observe that

$$\log(A + x^\alpha) - \alpha \log(B + x)$$

is concave up, for positive A and B, to conclude that the maximum occurs at an
endpoint, and the upper endpoint $x = 3^{-m}$ yields a value of V that is too small.

For fixed m, we can study the behavior of $\log V - \alpha \log L$ as a function of a_m by
replacing a_m by a real variable y and differentiating. We find that the derivative is
negative for all $y > 0$ provided

$$4^m - 1 \leq \alpha 9^m \left(1 - \frac{4}{5 \cdot 6^m} - \left(\frac{2}{3}\right)^m\right),$$

and this holds for all $m \geq 2$. Thus we can assume $a_m = 0$ if $m \geq 2$. But with $x = 0$
and $a_m = 0$ we have

$$\log V - \alpha \log L = \log \left(1 - \frac{4}{5 \cdot 6^m} - \left(\frac{2}{3}\right)^m\right) - \alpha \log \left(\frac{4^m - 1}{3^{m-1}}\right) - \alpha \log \frac{6}{\sqrt{3}},$$

and it is not hard to see that this is decreasing for $m \geq 2$. Finally, we have only
to search for the maximum among the 6 possibilities, $m = 1$ and $0 \leq a_1 \leq 5$, or
$m = 2$ and $a_2 = 0$, and $m = 1$ and $a_1 = 5$ wins.   Q.E.D

For small values of L, it is clear that the isoperimetric inequality (4.11) cannot
be improved, since we have equality for a sequence of values approaching zero. This

does not mean, however, that we have a complete description of the isoperimetric profile function for small L. It is clear that there is some value $L_0 > 28/\sqrt{3}$ such that for $28/\sqrt{3} \le 3^k L \le L_0$ we have $h(L) = 6^{-k} h(3^k L)$, with the optimal region being similar to the big region that is optimal for $3^k L$. However, we must have $L_0 < 50/\sqrt{3}$, since $h(50/\sqrt{3}) = 34/45 = .755\ldots$ ($m = 3$, $a_3 = 27$) while $h(24/\sqrt{3}) = 11/30 = .366\ldots$ ($m = 1$, $a_1 = 3$) and $h(26/\sqrt{3}) = 19/45 = .422\ldots$ ($m = 1$, $a_1 = 4$). Thus the union of scaled down images of the two domains ($m = 1$, $a_1 = 3$ and $m = 1$, $a_1 = 4$) has larger measure than the single domain ($m = 3$, $a_3 = 27$), with the same length perimeter. The description of $h(L)$ and the corresponding optimal domains for the values $L_0 < 3^k L < 54/\sqrt{3}$ is likely to be quite involved.

For large values of L, on the other hand, (4.11) is completely trivial ($V \to 1$ as $L \to \infty$), so we need an entirely different sort of estimate. To see what this should be like, we observe that the main terms in (4.8) and (4.9) yield $L \approx (4/3)^m 18/\sqrt{3}$ and $V \approx 1 - (2/3)^m$, so $V \approx 1 - (L\sqrt{3}/18)^{-\gamma}$, where

$$(4.14) \qquad \gamma = \log(3/2)/\log(4/3) \approx 1.409421.$$

To describe a more precise estimate we need one more quantity,

$$(4.15) \qquad \delta = \log 3/\log(4/3) \approx 3.8188417.$$

Theorem 4.4. *For any $c > \gamma - 4/5$ and for large L we have*

$$(4.16) \qquad V \le 1 - (L\sqrt{3}/18)^{-\gamma} + c(L\sqrt{3}/18)^{-\gamma-1-\delta}.$$

Proof. Write $a_m + 3^m x = y$ in (4.8), so

$$(4.17) \qquad 0 \le y \le 4^m + 2,$$

and (4.8) can be written

$$(4.18) \qquad L\sqrt{3}/18 = \left(\frac{4}{3}\right)^m - \left(\frac{1}{3}\right)^m + \frac{y}{3^{m+1}}.$$

We use (4.13) to obtain an upper bound for V from (4.9) in terms of y:

$$(4.19) \qquad V \le 1 - \left(\frac{2}{3}\right)^m + \left(\frac{y}{3} - \frac{4}{5}\right)\left(\frac{1}{6}\right)^m.$$

Indeed, to establish (4.19) we start with

$$(4.20) \qquad V \le 1 - \left(\frac{2}{3}\right)^m - \left(\frac{4}{5}\right)\left(\frac{1}{6}\right)^m + \frac{1}{3}\left(a_m\left(\frac{1}{6}\right)^m + x^\alpha\right)$$

from (4.9) and (4.13), and then show that

$$(4.21) \qquad a_m\left(\frac{1}{6}\right)^m + x^\alpha \le y\left(\frac{1}{6}\right)^m,$$

which is equivalent to $x^\alpha \le x\left(\frac{1}{2}\right)^m$ by the definition of y; and this holds because $x \le \left(\frac{1}{3}\right)^m$.

To simplify the notation we write $u = \left(\frac{4}{3}\right)^m$, and note that $u^{-\gamma} = \left(\frac{2}{3}\right)^m$, $u^{-\delta} = \left(\frac{1}{3}\right)^m$ and $u^{-\gamma-1-\delta} = \left(\frac{1}{6}\right)^m$. In view of (4.18) and (4.19) it suffices to prove that

$$(4.22) \qquad \begin{aligned} 0 \le\ & u^{-\gamma} - (u - u^{-\delta} + (y/3)u^{-\delta})^{-\gamma} \\ & + c(u - u^{-\delta} + (y/3)u^{-\delta})^{-\gamma-\delta-1} + \left(\frac{4}{5} - \frac{y}{3}\right)u^{-\gamma-\delta-1} \end{aligned}$$

for large u. Now the y derivative of the right side of (4.22) is

$$(\gamma/3)u^{-\delta}(u - u^{-\delta} + (y/3)u^{-\delta})^{-\gamma-1} - (1/3)u^{-\gamma-1-\delta}$$
$$- (\gamma/3)c(\gamma + \delta + 1)u^{-\delta}(u - u^{-\delta} + (y/3)u^{-\delta})^{-\gamma-\delta-2}$$

and the dominant term is $((\gamma - 1)/3)u^{-\gamma-\delta-1}$. So for large u, the right side of (4.22) is an increasing function of y. Thus it suffices to prove (4.22) for $y = 0$. But then the right side is

$$(c - \gamma + 4/5)u^{-\gamma-\delta-1} + O(u^{-\gamma-\delta-2}),$$

which is positive for large u because of our assumption on c.   Q.E.D

It is clear from the proof that the estimate (4.16) is sharp, although it would be possible to take $c = \gamma - 4/5$ and add a term of higher order.

5. Isoperimetric estimates on the pentagasket

The pentagasket (PG_5) has dimension

$$\alpha = \log 5/2 \log \tau \approx 1.672276$$

and connectivity dimension

$$\beta = \log(1 + \sqrt{3})/2 \log \tau \approx 1.0442933,$$

where $\tau = (1 + \sqrt{5})/2$ is the golden ratio. We let L denote the $\mathcal{H}_\beta$ measure of a curve, normalized so that the minimum path joining consecutive vertices of the pentagon (called Γ_1) has $L = 1$. It follows from (2.16) that $L = (1 + \sqrt{3})/2$ for the minimum path joining non–consecutive vertices, which we call Γ_2. We need to compute the $\mathcal{H}_\alpha$ measure (which we denote by V, and normalize so the total measure of the pentagasket is 1) of the two regions that Γ_2 cuts off, called A (the smaller) and B.

Lemma 5.1. *The A and B regions cut by Γ_2 have $V = 5/17$ and $V = 12/17$.*

Proof. As shown in Figure 5.1, A can be dissected into 2 A regions contracted once, 2 B regions contracted twice, and 3 pentagons contracted twice. This leads to the identity $V = 2V/5 + 2(1 - V)/25 + 3/25$ for the A region.   Q.E.D

We call a region *big* if it contains the big deleted pentagon. We compute the isoperimetric profile function for big regions first. The smallest big region is shown in Figure 5.2. Its boundary passes through the five junction points, and consists of 5 once contracted Γ_2 curves. This region can be dissected into 10 twice contracted A regions. The scaling factor for L is $\tau^{-2\beta} = (1+\sqrt{3})^{-1}$. Thus we find $L = 5/(1+\sqrt{3})$ and $V = 2/17$ for this region. We build up larger big regions by adding on certain numbers of deleted pentagons of order $1, 2, \ldots$. However, there are two distinct ways we can do these additions, which we call *major* and *minor* additions, as shown in Figure 5.3. A major addition of order k replaces 1 Γ_1 curve of order k by 3 Γ_1 and 2 Γ_2 curves of order $k+1$, for a net gain of 3 $(1+\sqrt{3})^{-k-1}$ in L. A minor addition of order k replaces 1 Γ_2 curve of order k by 2 Γ_1 and 2 Γ_2 curves of order $k+1$, for a net gain of $(1+\sqrt{3})^{-k-1}$ in L. The change in measure is $(4/17)5^{-k}$ for both types of additions, but for different reasons. For a major addition we replace 2 A regions of order $k+1$ by 2 B regions of order $k+1$ and 6 A regions of order $k+2$.
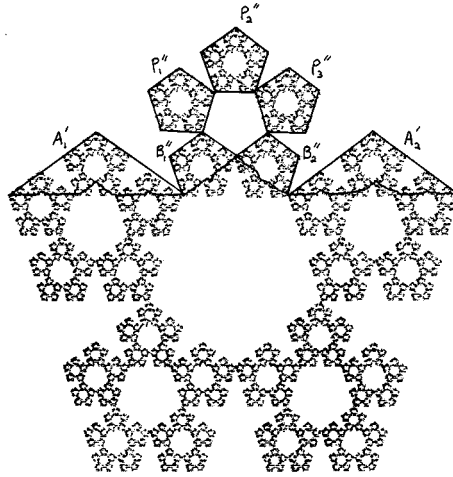
FIGURE 5.1. The small region A dissected into $A'_1 \cup A'_2 \cup B''_1 \cup B''_2 \cup P''_1 \cup P''_2 \cup P''_2$, where A'_1 and A'_2 are small regions contracted once, B''_1 and B''_2 are large regions contracted twice, and P''_1, P''_2, P''_3 are pentagons contracted twice.
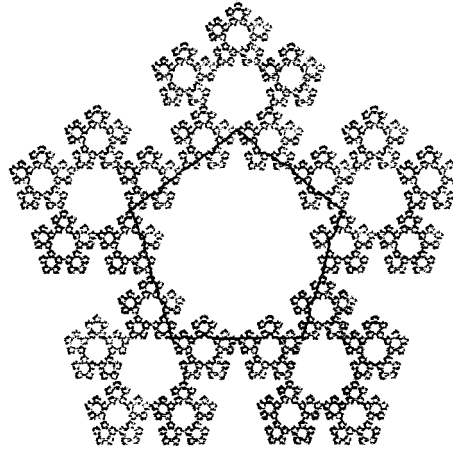

FIGURE 5.2. The smallest big region, which consists of 10 twice contracted A regions. It is bounded by 5 once contracted Γ_1 curves, and equivalently, 10 twice contracted Γ_2 curves.

For a minor addition we replace 1 A region of order k by 1 pentagon of order k, 2 B regions of order $k+1$ and 4 A regions of order $k+2$.

Let a_k denote the number of major additions and b_k the number of minor additions of order k. Then

$$(5.1) \qquad L = \frac{5}{(1+\sqrt{3})} + \sum_{k=1}^{\infty} \frac{3a_k + b_k}{(1+\sqrt{3})^{k+1}}$$
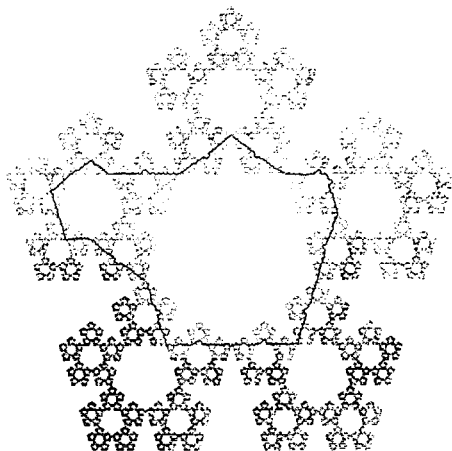
FIGURE 5.3. The smallest big region with one major addition of
order 1 and one minor addition of order 2.

and

$$(5.2) \qquad V = \frac{2}{17} + \sum_{k=1}^{\infty} \frac{4(a_k + b_k)}{17 \cdot 5^k}.$$

Our next task is to find recursive bounds on the values of a_k and b_k. To do this we
need to describe the boundary of the region after $k-1$ steps of the construction. It
will consist of a certain number of Γ_1 and Γ_2 curves of order k, but these can occur
in two different orientations, which we call *standard* and *reversed*. The standard
orientation has the smaller part of the order k pentagon inside the region, and the
reversed orientation has the larger part inside. Additions are possible only from
the standard orientation; major additions from Γ_1 curves and minor additions from
Γ_2 curves. We denote by c_k and d_k the number of Γ_1 and Γ_2 curves in standard
orientation, and by $\widetilde{c}_k$ and $\widetilde{d}_k$ the number in reversed orientation. Then we have

$$(5.3) \qquad 0 \le a_k \le c_k,$$

$$(5.4) \qquad 0 \le b_k \le d_k,$$

and any choices in these ranges are possible. We observe that either type of addition
adds the Γ_1 curves in standard orientation and the Γ_2 curves in reversed orientation.
If we do not perform a major addition on a Γ_1 curve of order k, it subdivides into
2 Γ_2 curves of the same orientation of order $k+1$. If we do not perform a minor
addition on a Γ_2 curve of order k, it subdivides into 2 Γ_2 curves of the same
orientation and 1 Γ_1 curve of the opposite orientation of order $k+1$. Thus we
obtain the recursion formulas

$$(5.5) \qquad \begin{cases} c_{k+1} = 3a_k + 2b_k + \widetilde{d}_k, \\ d_{k+1} = -2a_k - 2b_k + 2c_k + 2d_k, \\ \widetilde{c}_{k+1} = -b_k + d_k, \\ \widetilde{d}_{k+1} = 2a_k + 2b_k + 2\widetilde{c}_k + 2\widetilde{d}_k. \end{cases}$$

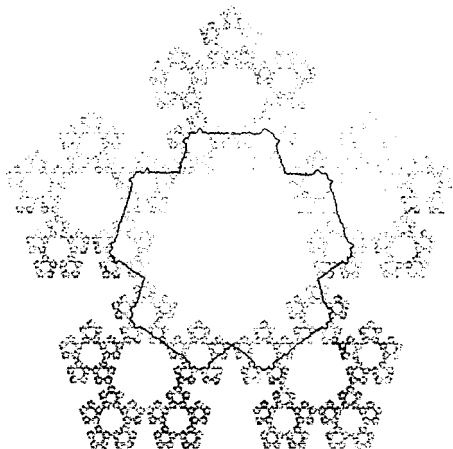For $k = 1$ we have $c_1 = 5$ and $d_1 = \widetilde{c}_1 = \widetilde{d}_1 = 0$.

FIGURE 5.4. The extremal region for (5.6), with $b_2 = 10$ and all other a_k and b_k zero.

Lemma 5.2. *The maximum value for a_k is $(5 + 10 \cdot 4^{k-1})/3$ with $b_k = 0$. This choice leads to $V = 1$.*

Proof. Choosing $a_k = c_k$ for all k leads to $b_k = d_k = \widetilde{c}_k = 0$ for all k. Then (5.5) reduces to

$$\begin{pmatrix} c_{k+1} \\ \widetilde{d}_{k+1} \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} c_k \\ \widetilde{d}_k \end{pmatrix},$$

from which we conclude that

$$\begin{pmatrix} c_k \\ d_k \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}^{k-1} \begin{pmatrix} 5 \\ 0 \end{pmatrix};$$

hence $c_k = (5 + 10 \cdot 4^{k-1})/3$ by linear algebra. From (5.2) we obtain $V = 1$ (of course $L = \infty$ from (5.1)).   Q.E.D

Because of the complexity of (5.5) we have not been able to determine the isoperimetic profile function exactly. However, we can still obtain sharp isoperimetric estimates. First we consider the estimate

$$(5.6) \qquad\qquad\qquad\qquad V \leq cL^\gamma$$

for $\gamma = \alpha/\beta = \log 5/\log(1 + \sqrt{3}) \approx 1.601347$.

Theorem 5.3. *The best constant in (5.6) is $c = \frac{8}{5 \cdot 17} \cdot \left(\frac{5}{1+\sqrt{3}} + \frac{10}{(1+\sqrt{3})^3} \right)^{-\gamma} \approx .0244481$ with the extremal region shown in Figure 5.4 ($b_2 = 10$, and all other a_k and b_k are zero). Equality holds in (5.6) if and only if the region is similar to this one.*

Proof. We consider first big regions, where L and V are given by (5.1) and (5.2), and we seek to maximize $\log V - \gamma \log L$. If we fix all variables except one a_k or b_k and replace it with a real variable x, then we have a function of the form

$$(5.7) \qquad\qquad \log(V_0 + vx) - \gamma \log(L_0 + \ell x),$$

and by calculus this will be decreasing in x for all $x > 0$ provided

$$(5.8) \qquad\qquad v/\ell \le \gamma V_0/L_0.$$

Now $v/\ell = \frac{4(1+\sqrt{3})}{51}\left(\frac{1+\sqrt{3}}{5}\right)^k$ for a_k and $\frac{4(1+\sqrt{3})}{17}\left(\frac{1+\sqrt{3}}{5}\right)^k$ for b_k, and these values get very small for large values of k. The value of $\gamma V_0/L_0$ depends on the other variables, and could also be small if L_0 is large. However, it is clear that the maximum in (5.6) will not occur for large L_0. For the region in Figure 5.4, $\gamma V_0/L_0 \approx$ .146, and the value will not be substantially different for any region that could plausibly come near the maximum. This eliminates a_k for $k \ge 2$ and b_k for $k \ge 3$. Thus we need only search through the possibilities

$$L = \frac{5}{(1+\sqrt{3})} + \frac{3a_1}{(1+\sqrt{3})^2} + \frac{b_2}{(1+\sqrt{3})^3},$$

$$V = \frac{2}{17} + \frac{4a_1}{17\cdot 5} + \frac{4b_2}{17\cdot 25}$$

with $0 \le a_1 \le 5$, $0 \le b_2 \le 10 - 2a_1$. (Rather than a brute force numerical calculation, we use calculus to show first that $\log V - \gamma \log L$ is increasing with b_2, so we must take $b_2 = 10 - a$, and then $L = \frac{5}{1+\sqrt{3}} + \frac{10}{(1+\sqrt{3})^3} + \frac{(1+3\sqrt{3})a_1}{(1+\sqrt{3})^3}$, $V = \frac{8}{5\cdot 17} + \frac{12a_1}{25\cdot 17}$; but then (5.8) holds, so the maximum is at $a_1 = 0$.) This proves (5.6) for the claimed value of c for big regions, with equality holding only for the region in Figure 5.4.

But (5.6) scales appropriately under the similarities, so it continues to hold for regions that are similar to a big region. Because $\gamma > 1$, (5.6) is preserved under disjoint unions of regions, with equality ruled out for non–trivial unions. Thus it holds for all regions.   Q.E.D

We discuss briefly the relationship between L and V as $V \to 1$ and $L \to \infty$. In order to get V close to 1 we have to take the maximum value for a_k described in Lemma 5.2 for all small values of k. To be specific, suppose we make this choice for all $k \le m$ and then take $a_k = b_k = 0$ for $k > m$. We obtain a region with

$$V = \frac{2}{17} + \sum_{k=1}^{m} \frac{20}{3\cdot 17\cdot 5^k}(1 + 2\cdot 4^{k-1}) = 1 - \frac{5}{51}\left(\frac{1}{5}\right)^m - \frac{40}{51}\left(\frac{4}{5}\right)^m$$

and

$$L = \frac{5}{1+\sqrt{3}} + \sum_{k=1}^{m} \frac{5 + 10\cdot 4^{k-1}}{(1+\sqrt{3})^{k+1}}$$

$$= \frac{5}{2(3-\sqrt{3})}\left(\frac{4}{1+\sqrt{3}}\right)^{m+1} + \left(\frac{5(1+\sqrt{3})}{\sqrt{3}} - \frac{5}{6-2\sqrt{3}}\right) - \frac{5}{\sqrt{3}(1+\sqrt{3})^{m+1}}.$$

Ignoring the lower order terms, we have $V \sim 1 - c_1(4/5)^m$ and $L \sim c_2(4/(1+\sqrt{3}))^m$, so

$$(5.9) \qquad\qquad V \sim 1 - cL^{-\delta}$$

for

$$(5.10) \qquad\qquad \delta = \frac{\log 5/4}{\log 4/(1+\sqrt{3})} \approx .5853071.$$

It would be possible to give a more precise estimate on the lines of Theorem 4.4, but we will not pursue the matter here.

6. Connectivity dimension of the octogasket

The octogasket is not point connected, so we cannot use the method of Section 2 to compute its connectivity dimension. The i.f.s. generating the octogasket is $S_j z = e^{ij\pi/4} + \rho(z - e^{ij\pi/4})$, $0 \le j < \infty$, where $\rho = 1 - \sqrt{2}/2$. Here P is an octagon with vertices λ_j at the 8th roots of unity, and consecutive images $S_j P$ and $S_{j+1} P$ intersect along edges. Thus we do not have precise information about where a curve crosses from $S_j P$ to $S_{j+1} P$; there is a whole Cantor set of points on the common edge which belong to the octogasket. It would greatly simplify our arguments to assume that minimal paths can be chosen to cross these edges at the endpoints. In fact we will end up proving this by first constructing more complicated minimal paths that do not cross at the endpoints. In other words, there is an extraordinary lack of uniqueness among minimal paths. We denote by Γ_j a minimal path joining λ_k and λ_{k+j}, and let $\gamma_j = \mathcal{H}_\beta(\Gamma_j)$. By symmetry such paths are isometric for different k.

Next we describe two basic symmetry principles that we will use repeatedly, at different scales. The first, which we call *avoid extremities*, concerns Γ_2 curves, or more generally, any minimal curves joining points that lie on or interior to the straight line segment joining λ_k and λ_{k+2}. It says that we can always replace the minimal curve by another minimal curve with no points exterior to the line segment. Figure 6.1 shows such a line segment. It is clear that the reflection in this line segment maps the exterior region into the octogasket. If Γ denotes a minimal curve, write $\Gamma = \Gamma' \cup \Gamma''$, where Γ'' denotes the portion of Γ lying in the exterior region. Let Γ''' denote the reflected image of Γ''. Then $\Gamma' \cup \Gamma'''$ is still a curve lying in the octogasket, joining the same two points, with the same dimension
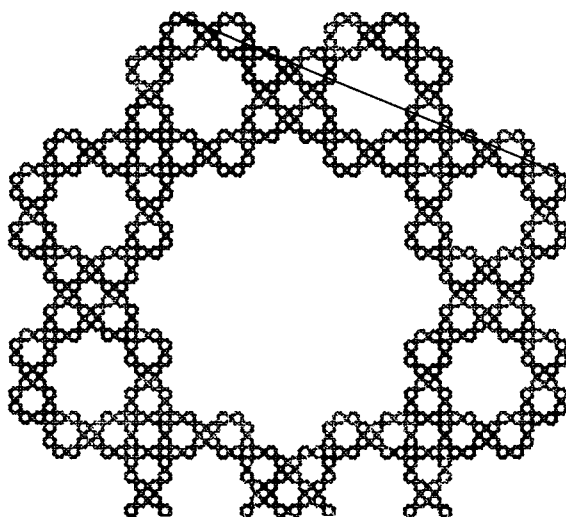

FIGURE 6.1. A straight line segment joining λ_0 and λ_2. Note that reflection in the line segment maps the exterior region into the octogasket.
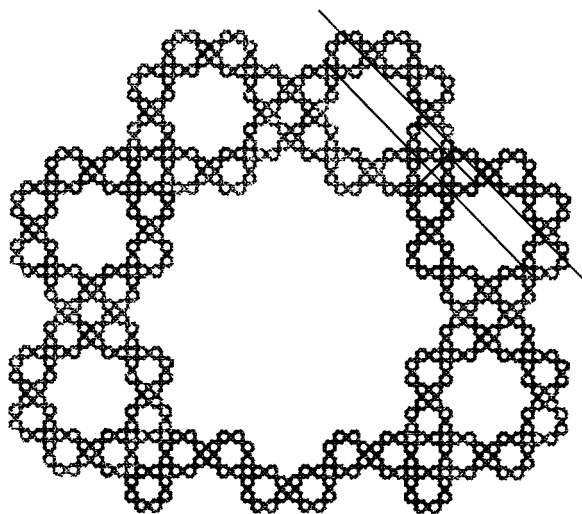
FIGURE 6.2. The perpendicular to the common edge at $S_0 P$ and $S_1 P$, at the outer corner. Also indicated are the perpendiculars at the midpoint, and the midpoint of the exterior half of the edge.

and measure, and with no points in the exterior region. This proves the avoid extremities principle.

The second principle we call *cross at the corner*. Consider a minimal curve joining points in $S_k P$ and $S_{k+1} P$, and suppose these points lie on or exterior to the perpendicular to the common edge at the outer corner (see Figure 6.2). Then we can replace the minimal curve by another minimal curve that passes through the corner point. To prove this we first consider the perpendicular bisector to the edge. We can reflect to the exterior any portion of the minimal curve that strays to the interior of this bisector. Thus, without loss of generality, the minimal curve lies on the exterior side of the bisector, and so must cross the common edge along the exterior half. Draw the perpendicular bisector to this half, and repeat the argument to get the minimal curve exterior to this. By iterating the argument, we force the minimal curve to cross at the corner.

Theorem 6.1. *The connectivity dimension β of the octogasket is given by*

$$(6.1) \qquad\qquad \rho^{-\beta} = (3 + \sqrt{17})/2 \qquad (\beta \approx 1.0344065),$$

and the regularity estimate (2.2) holds. Also

$$(6.2) \qquad\qquad\qquad \gamma_1 = 2\rho^\beta \gamma_2,$$

$$(6.3) \qquad\qquad\qquad \gamma_3 = (1 + \rho^\beta)\gamma_2,$$

and

$$(6.4) \qquad\qquad\qquad \gamma_4 = (1 + 2\rho^\beta)\gamma_2.$$

Proof. To prove (6.2) we apply the cross at the corner principle to a Γ_1 curve to replace it by the union of 2 once contracted Γ_2 curves.
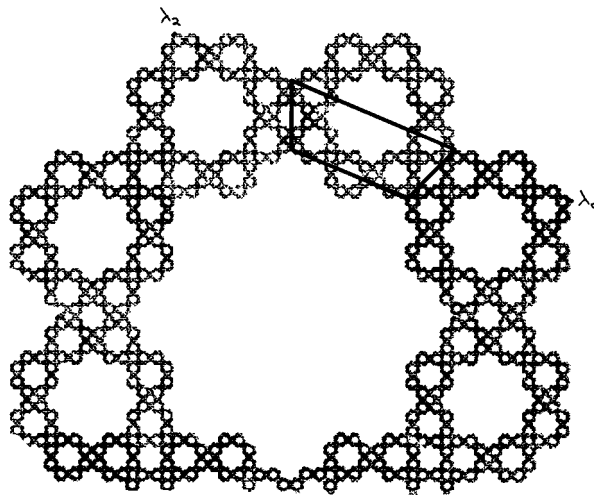
FIGURE 6.3. The Γ_2 curve joining λ_0 and λ_2, during its sojourn in $S_1 P$, must lie in the trapezoidal region, first because it lies interior to the long base, then because it lies exterior to the short base. By iterating this argument, it must pass through z_1.

Next we analyze a Γ_2 curve. We claim that the avoid extremities principle can be applied iteratively to force Γ_2 to pass through the point z_{k+1}, which is the fixed point of the map $S_{k+1} S_{k+5}$ (so $z_k = \lim_{N \to \infty} (S_{k+1} S_{k+5})^N 1$). Applied once, the principle forces Γ_2 to lie in the interior half of $S_{k+1} P$. But then consider the straight line segment in $S_{k+1} P$ joining the two interior vertices where $S_{k+1} P$ intersects $S_k P$. Applying the avoid extremities principle forces Γ_2 to stay exterior to this line segment. In particular, it passes through the exterior half of $S_{k+1} S_{k+5} P$. By iterating this argument, flipping back and forth between interior and exterior, we force Γ_2 to pass through z_{k+1}. See Figure 6.3. In fact, the same argument shows that any minimal curve that passes through $S_{k+1} P$ may be assumed to pass through z_{k+1}.

Thus Γ_2 consists of two halves (which may be assumed to be isometric) joining λ_k to z_{k+1} and joining z_{k+1} to λ_{k+2}. So it suffices to understand the first half. Since it begins in $S_k S_k P$ and passes through $S_k S_{k+1} P$ on its way to $S_k S_{k+2} P$, it can be forced to pass through the point $S_k z_{k+1}$ (see Figure 6.4). Now we observe an interesting symmetry. Let M denote the midpoint of the common edge between $S_k P$ and $S_{k+1} P$ (M is not a point in the octogasket). The half–turn about M preserves the portion of the octogasket contained in $S_k P \cup S_{k+1} P$, and it interchanges the points z_{k+1} and $S_k z_{k+1}$. So the minimal curves joining these two points come in pairs with respect to this symmetry. Without loss of generality we consider a minimal curve that crosses the common edge between $S_k P$ and $S_{k+1} P$ along the interior half. The same reasoning as before then forces the curve to pass through the points marked q_3 and q_2 in Figure 6.4. We label z_{k+1} by q_4 and $S_{k+1} z_{k+1}$ by q_1, and break up the half of Γ_2 into $A \cup B \cup C \cup D$, where A joins z_k to q_1, B joins q_1 to q_2, C joins q_2 to q_3, and D joins q_3 to q_4. If we denote the $\mathcal{H}_\beta$ measure of
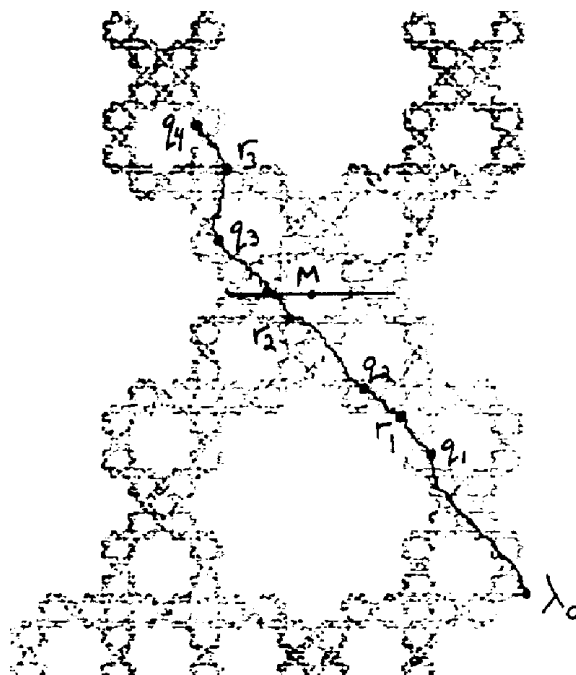
FIGURE 6.4. A half Γ_2 curve joining λ_0 and $z_1 = q_4$. The curve may be forced to pass through the points q_1, q_2, q_3, q_4 and r_1, r_2. The midpoint M of the common edge of $S_0 P$ and $S_1 P$ is a point of symmetry for q_1 and q_4.

those curves by lower case letters, we have the identity

$$(6.5) \qquad\qquad \gamma_2/2 = a + b + c + d.$$

We also note that the cross at the corner principle forces the curve to pass through the points labeled r_1 and r_2 in Figure 6.4, which are midpoints of segments B and C.

Now A is a once contracted half Γ_2 curve, so $a = \rho^\beta \gamma_2/2$. Also B is the union of two halves, each of which is a twice contracted half Γ_2 curve, so $b = \rho^{2\beta}\gamma_2$. D is just a once contracted image of $B \cup C \cup D$, so $d = \rho^\beta(b + c + d)$, whence $d = \rho^\beta(1 - \rho^\beta)^{-1}(b + c)$. Half of the C curve, joining q_2 to r_2, is a once contracted image of a minimal curve joining q_1 to the point labeled r_3 in Figure 6.4, which is the union of B, C and a minimal curve joining q_3 to r_3. Now a minimal curve joining q_3 to r_3 is a twice contracted image of a minimal curve joining λ_{k-1} and z_{k+1}, which is half of a Γ_4 curve, as we will demonstrate in the course of proving (6.4). Thus we have $c = 2\rho^\beta(b + c + \rho^{2\beta}\gamma_4/2)$; hence

$$c = (1 - 2\rho^\beta)^{-1}\rho^\beta(2b + \rho^{2\beta}\gamma_4).$$
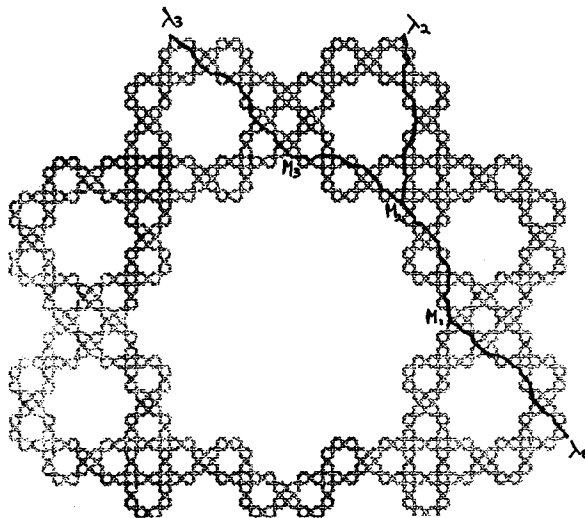
FIGURE 6.5. A simpler choice for minimal Γ_2 and Γ_3 paths.

Using these computations, and (6.4), and writing $\rho^\beta = x$, we have

$$a = \frac{x}{2}\ \gamma_2,$$
$$b = x^2\gamma_2,$$
$$c = \frac{3x^3 + 2x^4}{1 - 2x}\ \gamma_2,$$
$$d = \frac{x^3 + x^4 + 2x^5}{(1 - x)(1 - 2x)}\ \gamma_2.$$

When we substitute these into (6.5) we can cancel γ_2 and obtain an equation involving x alone, which simplifies to

(6.6) $$4x^4 + 4x^3 - 3x^2 + 4x - 1 = 0.$$

But this equation factors into $(2x^2 + 3x - 1)(2x^2 - x + 1) = 0$, and the second factor has no positive zero. Thus $\rho^{-\beta} = x^{-1}$ satisfies

(6.7) $$\rho^{-2\beta} - 3\rho^{-\beta} - 2 = 0,$$

which proves (6.1).

We turn now to the analysis of the Γ_3 and Γ_4 curves. We apply the cross at the corner principle to force Γ_3 to pass through the interior corner M_2 of the common edge between $S_{k+1}P$ and $S_{k+2}P$, which then forms a symmetric midpoint for Γ_3 (see Figure 6.5). Looking at the first half of Γ_3, just as for Γ_2, we can force it to pass through the point z_{k+1}. The portion from λ_k to z_{k+1} is a half of Γ_2, and the portion from z_{k+1} to the midpoint M_1 is a once contracted image of half of Γ_2. This proves (6.3).

Now consider a Γ_4 curve. It is clear that we may go in either a clockwise or a counterclockwise direction, and we may assume the latter. By the same reasoning as before we can force Γ_4 to pass through z_{k+1}, M_1 and z_{k+2}, and now z_{k+2} serves
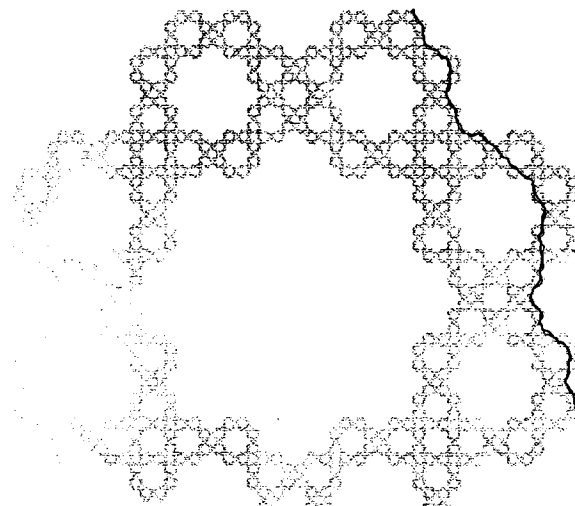
FIGURE 6.6. A difference choice of Γ_2 path that crosses at the exterior corners of the edges.

as a midpoint for Γ_4. Thus a half Γ_4 curve is made up of a half Γ_2 curve and 2 once contracted half Γ_2 curves, which yields (6.4).

The minimal curves Γ_2, Γ_3 and Γ_4 constructed above form a family of self–similar sets with irreducible matrix M_β, so the values γ_2, γ_3, γ_4 are finite and positive, and the regularity estimate follows as in Theorem 2.5.   Q.E.D

After the fact, we can describe a much simpler family of minimal paths. First we claim that there exists a Γ_2 path that passes through the points M_1 and M_2 in Figure 6.5 (the interior endpoints of the edges). The reason is that we can take scaled versions of any Γ_2 path to join M_1 and M_2, and scaled versions of any Γ_3 path to join λ_0 to M_1 and M_2 to λ_2. The resulting path would have $\mathcal{H}_\beta$ measure equal to $\rho^\beta(\gamma_2 + 2\gamma_3)$, which is equal to γ_2 by (6.3) and (6.7). So it, too, is a minimal path. Similarly, we can find a Γ_3 path that passes through M_1, M_2, and M_3 (see Figure 6.5), since such a path would have $\mathcal{H}_\beta$ measure equal to $\rho^\beta(2\gamma_2 + 2\gamma_3)$, which equals γ_3 by (6.3) and (6.4).

It then follows that we can take minimal Γ_2 and Γ_3 paths that form a family of self–similar sets by iterating these choices, and this yields the matrix

$$M_\beta = \rho^\beta \begin{pmatrix} 2 & 2 \\ 2 & 1 \end{pmatrix},$$

which immediately yields (6.7), hence (6.1). Surprisingly, we can also construct minimal paths that cross the edges at the exterior corners (see Figure 6.6). This means a Γ_2 path that consists of 2 scaled copies of Γ_2 and 1 scaled copy of Γ_4, giving a total $\mathcal{H}_\beta$ measure of $\rho^\beta(2\gamma_2 + \gamma_4)$, which also equals γ_2 by (2.4) and (2.7).

7. Isoperimetric estimates on the octogasket

The octogasket has dimension

$$\alpha = \log 8 / \log(1 - \sqrt{2}/2)^{-1} \approx 1.6934291$$

and connectivity dimension

$$\beta = \log((3+\sqrt{17})/2)/\log(1-\sqrt{2}/2)^{-1} \approx 1.0344065;$$

hence

$$(7.1) \qquad \gamma = \alpha/\beta = \log 8/\log((3+\sqrt{17})/2) \approx 1.6371022.$$

We normalize the $\mathcal{H}_\beta$ measure so that $L = 1$ for Γ_2. Then Theorem 6.1 gives the values $2x$, $1+x$ and $1+2x$ for L for the paths Γ_1, Γ_3 and Γ_4, where

$$(7.2) \qquad x = (1-\sqrt{2}/2)^\beta \approx .2807764$$

(we have $2x^2 + 3x - 1 = 0$ by (6.7)). Our goal is to find the best constant in the isoperimetric estimate

$$(7.3) \qquad V \leq cL^\gamma$$

and to find the asymptotic growth of L as $V \to 1$.

One of the complications that arises with the octogasket is the severe lack of uniqueness of minimum paths. However, it is not difficult to find the ones that maximize and minimize the $\mathcal{H}_\alpha$ measure of the regions cut off. We write Γ_2^+ and Γ_2^- for these choices of Γ_2 path, and V_2^+ and V_2^- for the measure of the region lying exterior to $\Gamma_2^\pm$, and similarly for $\Gamma_j^\pm$ and $V_j^\pm$ for other values of j (for $j = 4$ we define the exterior direction so that V_4^+ is less than half the total measure). It turns out that the $\Gamma_j^\pm$ paths are unique, and they are determined by the following decomposition rules:

$$(7.4) \qquad \begin{cases} \Gamma_1^- \to 2\Gamma_2^-, \\ \Gamma_1^+ \to 2\Gamma_2^+, \\ \Gamma_2^- \to 2\Gamma_2^- + \Gamma_4^-, \\ \Gamma_2^+ \to \Gamma_2^- + 2\Gamma_3^+, \\ \Gamma_3^- \to 2\Gamma_2^- + 2\Gamma_3^+, \\ \Gamma_3^+ \to 2\Gamma_2^- + 2\Gamma_3^+, \\ \Gamma_4^- \to 2\Gamma_2^- + \Gamma_2^+ + 2\Gamma_3^+, \\ \Gamma_4^+ \to 3\Gamma_2^- + 2\Gamma_3^+. \end{cases}$$

The meaning of these rules is that a $\Gamma_j^\pm$ path breaks up into a union of other such paths once contracted, and these are written to the right of the arrow (we use plus signs rather than union signs for notational clarity). These decompositions are illustrated in Figure 7.1, and the figures gives more precise information than (7.4). For example, both Γ_3^- and Γ_3^+ yield the same pieces in (7.4), but the ordering of the pieces is different.

Next, we use Figure 7.1 to write linear equations for the values of $V_j^\pm$. For example, V_2^+ is the measure of the region exterior to Γ_2^+, and this breaks up into three once contracted regions, two being exterior to Γ_3^+ paths and one being the complement of the exterior of a Γ_2^- path. Since contracting multiplies the measure
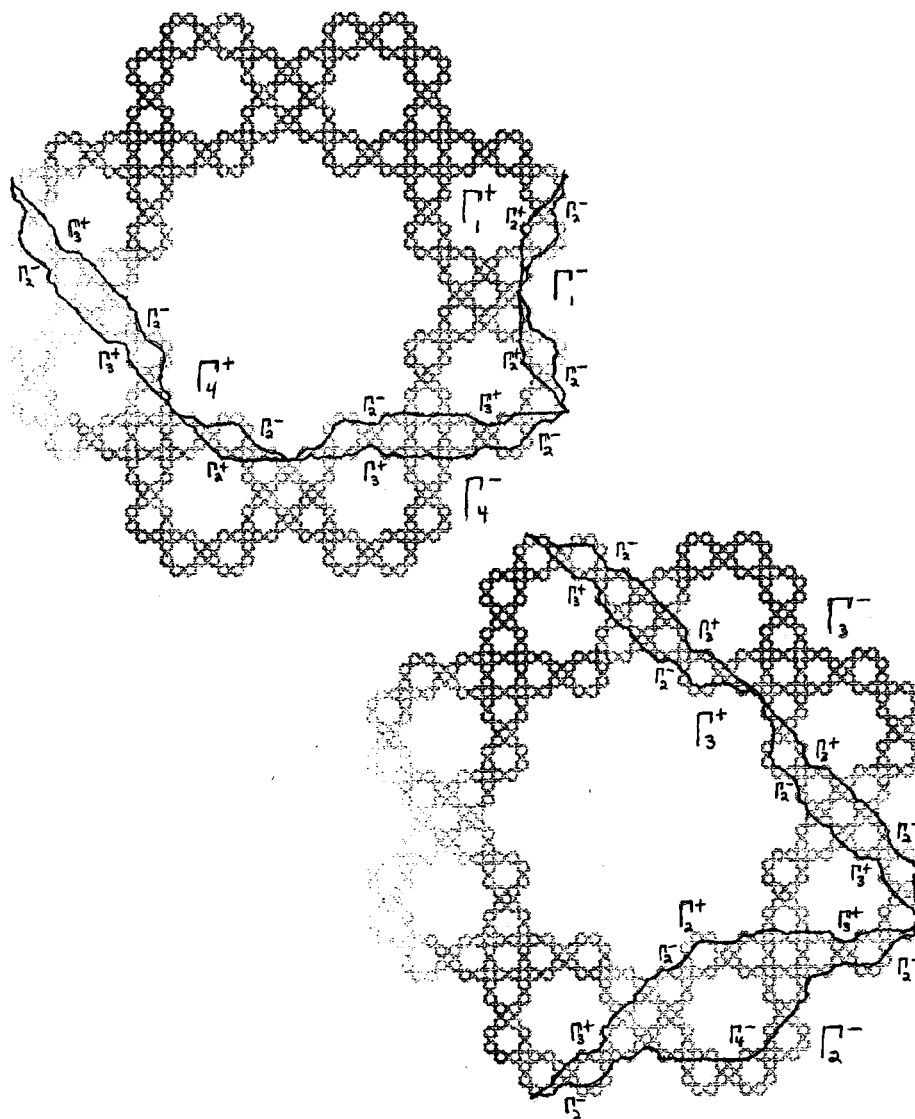
FIGURE 7.1. Paths of type $\Gamma_j^\pm$ for $j = 1, 2, 3, 4$, with the decomposition rules (7.4).

by 1/8, the equations are

$$(7.5) \quad \begin{cases} 8V_1^- = 2V_2^-, \\ 8V_1^+ = 2V_2^+, \\ 8V_2^- = 2V_2^- + V_4^-, \\ 8V_2^+ = (1 - V_2^-) + 2V_3^+, \\ 8V_3^- = 2V_2^- + 2(1 - V_3^+), \\ 8V_3^+ = 2(1 - V_2^-) + 2V_3^+, \\ 8V_4^- = 2V_2^- + (1 - V_2^+) + 2(1 - V_3^+), \\ 8V_4^+ = 3(1 - V_2^-) + 2V_3^+. \end{cases}$$
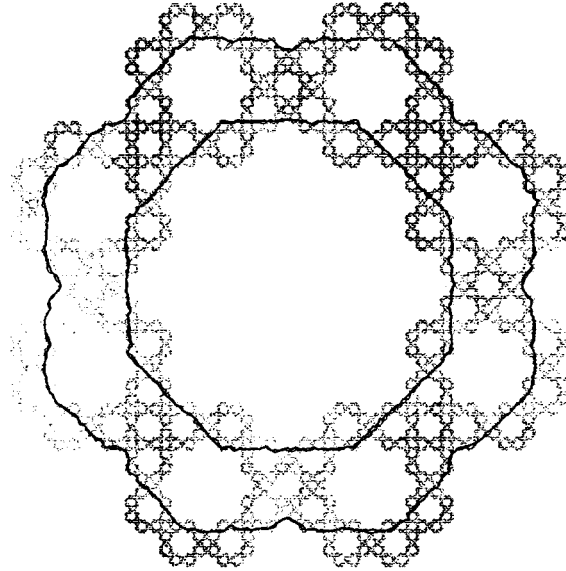
FIGURE 7.2. The boundaries of two big regions. The inner region is the big region with minimum L (boundary consisting of 8 Γ_2^+ paths). The outer region contains the 8 next largest deleted octogons, and is bounded by 8 Γ_4^- paths.

The solutions are given as follows:

$$(7.6) \quad \begin{cases} V_1^- = 51/4332, \\ V_1^+ = 215/4332, \\ V_2^- = 51/1083, \\ V_2^+ = 215/1083, \\ V_3^- = 395/2166, \\ V_3^+ = 344/1083, \\ V_4^- = 306/1083, \\ V_4^+ = 473/1083. \end{cases}$$

This is the information we need to compute the values of V and L for regions that have the potential to maximize V for the given L value.

We begin by considering big regions, which are defined to be regions containing the large deleted octagon. The minimum L value for an S_j region is $8x$, since the boundary path must pass through all 8 contracted octogons $S_j P$, and in each one the minimal length path is a contracted Γ_2 path. Clearly we maximize V for $L = 8x$ by taking Γ_2^+ paths (see Figure 7.2), which yields $V = 8 \cdot (1/8) V_2^+ = 215/1083$. Next we consider larger big regions by adding on some of the 8 next deleted octogons. The shortest path containing all 8 of these consists of 8 once contracted Γ_4 paths, and we maximize V for this value of $L = 8x(1 + 2x) = 8 - 16x \approx 3.5075776$ by taking Γ_4^- paths, which yields $V = (1 - V_4^-) = 777/683 \approx .7174515$ (see Figure 7.2). This region gives a value .0919517 for V/L^γ, and this turns out to be the
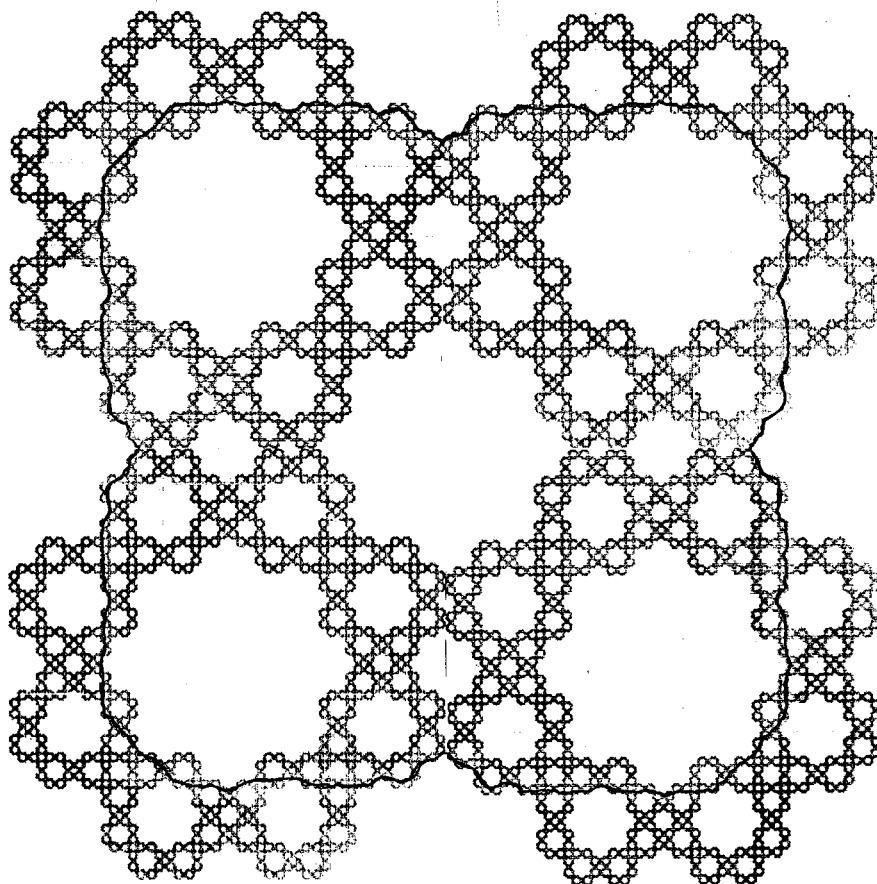
FIGURE 7.3. The region that attains the maximum constant in
(7.3), twice enlarged.

maximum value among big regions. We will not give the argument for this, since
we can get still larger values using regions that are not big.

   To see this, consider the region shown in Figure 7.3, which is the region with
minimal L containing the deleted octogon in each of the 4 twice contracted octogons
that meet along the common edge of 2 once contracted octogons. The boundary
consists of 16 Γ_2 paths and 8 Γ_3 paths (3 times contracted), which yields

$$L = x^3(24 + 8x) = x^2(4 + 12x).$$

Each of the 4 twice contracted octogons contains 2 entire octogons, 2 V_2^+ regions,
2 V_3^+ regions, and the complements of 2 V_2^- regions (all 3 times contracted) for a
total $V = (1/64)(2 - V_2^- + V_2^+ + V_3^+)$. Thus

$$V/L^\gamma = (2 - V_2^- + V_2^+ + V_3^+)/(4 + 12x)^\gamma \approx 2.4690674/(7.3693168)^\gamma \approx .0938568,$$

and this turns out to be the optional constant in (7.3). It might seem that we
could increase the ratio by taking a slight detour to enclose a larger V. But it turns
out that this forces an increase in L that more than compensates. We omit the
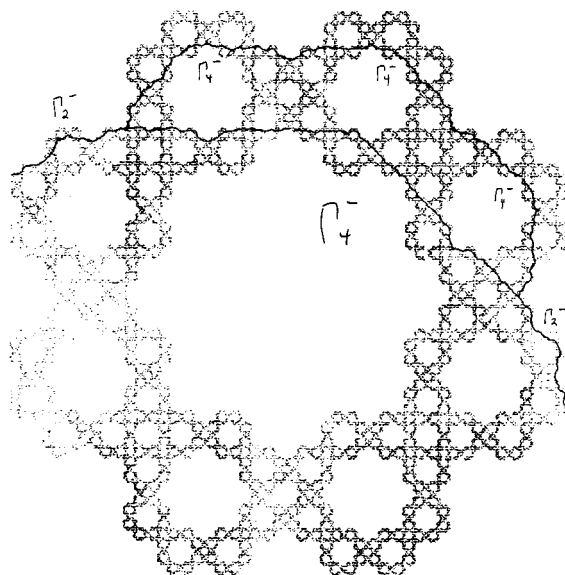
FIGURE 7.4. The process of enlargement, replacing a Γ_4^- path by
2 Γ_2^- and 3 Γ_4^- paths (contracted one more time).

details of the computation. Similarly, we obtain a smaller ratio if we take regions
containing 2 or 3 of the deleted octogons.

Next we consider a sequence of regions, each maximizing V for its value of L,
with $V \to 1$. We start with the larger region in Figure 7.2 bounded by 8 Γ_4^- paths.
At each stage we take each Γ_4^- path on the boundary and replace it with 2 Γ_2^- and
3 Γ_4^- paths (contracted one more time), as shown in Figure 7.4. Each such addition
increases L by $x^k(2-6x)$ and V by $(5V_4^- - 2V_2^-)/8^{k+1}$ going from stage k to stage
$k+1$. Also, each Γ_2^- path on the boundary is just subdivided into 2 Γ_2^- and 1 Γ_4^-
path (contracted one more time), which leaves L and V unchanged.

Let a_k and b_k denote the number of Γ_4^- and Γ_2^- paths on the boundary at stage
k. Then $a_1 = 8$, $b_1 = 0$, and the subdivision scheme yields the recursion relations

$$(7.7) \qquad \begin{cases} a_{k+1} = 3a_k + b_k, \\ b_{k+1} = 2a_k + 2b_k, \end{cases}$$

which are easily solved to yield

$$(7.8) \qquad \begin{cases} a_k = (4^{k+1} + 8)/3, \\ b_k = (4^{k+1} - 16)/3. \end{cases}$$

On the other hand, at stage k we have

$$(7.9) \qquad \begin{cases} L = x^k b_k + x^k(1+2x)a_k, \\ V = (1 - V_4^-) + \displaystyle\sum_{j=1}^{k-1} a_j(5V_4^- - 2V_2^-)/8^{j+1}. \end{cases}$$

Altogether we find that

$$(7.10) \qquad \begin{cases} L = c_1(4x)^k - c_2 x^k, \\ 1 - V = c_3/2^k + c_4/8^k, \end{cases}$$

for $c_1 = 8(1+x)/3$, $c_2 = 8(1-2x)/3$, $c_3 = 476/1083$, $c_4 = 544/1083$. Note that $4x > 1$, so $L \to \infty$ as $V \to 1$. Thus we have the asymptotic behavior

$$(7.11) \qquad V \sim 1 - cL^{-\delta} \quad \text{as} \quad V \to 1$$

for

$$(7.12) \qquad \delta = \log 2/\log(4x) \approx 5.970378.$$

8. Intrinsic metrics

Suppose we have a fractal F with connectivity dimension β satisfying the regularity condition (2.2). Then we can define a new metric

$$(8.1) \qquad d(x,y) = \inf\{\mathcal{H}_\beta(\Gamma)^{1/\beta} : \Gamma \text{ joins } x \text{ and } y\}$$

(the triangle inequality follows from $\beta \geq 1$) that is equivalent to the original metric. It can be argued that this metric is more intrinsic than the original one, which F inherited from the way it was embedded in $\mathbb{R}^n$. If we compute Hausdorff measures with respect to this new metric we can expect to get somewhat different values, but only up to a bounded multiple. However, it is not hard to see that we get exactly the same value for $\mathcal{H}_\beta$. Thus we do not get anything new by reiterating the procedure. Also, if F is self–similar, then $\mathcal{H}_\alpha$ on F is characterized up to a constant multiple by the self–similar identity

$$(8.2) \qquad \mu = \sum_{j=1}^m r_j^\alpha \mu \circ S_j^{-1},$$

so $\mathcal{H}_\alpha$ is only changed by a constant multiple. Thus all our isoperimetric estimates are essentially the same (except for a change in the constant) in the new metric.

We can describe this new metric completely for the Sierpinski gasket. The most convenient coordinate system to use is barycentric coordinates. Let us assume the original triangle P has vertices X, Y, Z (we do not have to assume they form an equilateral traingle). Then any point p inside P has barycentric coordinates (x,y,z) satisfying

$$(8.3) \qquad x + y + z = 1, \ x \geq 0, \ y \geq 0, \ z \geq 0,$$

where

$$(8.4) \qquad p = xX + yY + zZ.$$

Suppose we represent $x = .x_1 x_2 \ldots$, $y = .y_1 y_2 \ldots$ and $z = .z_1 z_2 \ldots$ in base 2. Then we can characterize the points in the Sierpinski gasket by the condition that for each k, exactly one of x_k, y_k, z_k is equal to 1. Another, more colorful way of saying this, is that the sum $x+y+z = .11\ldots$ holds with no carries. Points with ambiguous representations are junction points and satisfy the condition in two distinct ways (but not in two other ways), as for example $x = .100\ldots$, $y = .011\ldots$, $z = 0$ or $x = .011\ldots$, $y = .100\ldots$, $z = 0$.

Now consider a minimal path connecting X to a point p. If p lies in the Y triangle the path begins with a line segment joining X to the junction point $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$, while if p lies in the Z triangle we do the same with the junction point $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$. In both
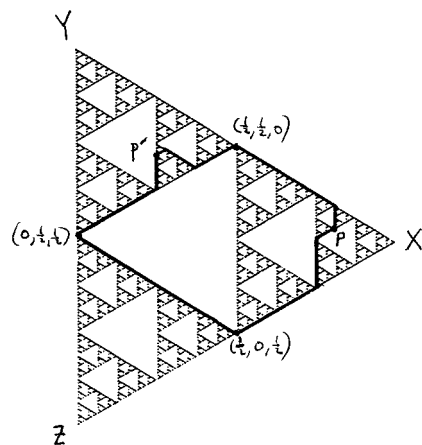
FIGURE 8.1. Two candidates for minimal path joining p (in the X triangle) and p' (in the Y triangle). One passes through the single junction point joining the two triangles, while the other goes through the other two junction points.

these cases we then go to the smaller triangle and iterate. If p lies in the X triangle we immediately go to this smaller triangle. In this way we always obtain a path consisting of a sequence of line segments. In the generic case this minimal path is unique and cannot be further extended (to form a minimal path). One might say that the typical minimal path is a maximal minimal path! Of course this will not be true if x, y or z is a dyadic rational. Note that in the cases of non–uniqueness, such as the point $\left(0, \frac{1}{2}, \frac{1}{2}\right)$, there are just two minimal paths. It is easy to see that the length of the minimal path is just

$$(8.5) \qquad |X - Y|y + |X - Z|z,$$

and a similar formula gives the lengths of minimal paths joining the other vertices Y and Z to an arbitrary point. In the case of an equilateral triangle (8.5) simplifies to

$$(8.6) \qquad c(1 - x),$$

where c is the length of a side, in view of (8.3).

Now suppose we have two distinct points $p = (x, y, z)$ and $p' = (x', y', z')$. Let k be the first integer for which $(x_k, y_k, z_k) \neq (x'_k, y'_k, z'_k)$. Then

$$(8.7) \qquad d(p, p') = 2^{1-k} d(\widetilde{p}, \widetilde{p}'),$$

where

$$(8.8) \qquad \widetilde{p} = (.x_k x_{k+1} \ldots , .y_k y_{k+1} \ldots , .z_k z_{k+1} \ldots),$$

and similarly for $\widetilde{p}'$. This it suffices to compute $d(p, p')$ when $k = 1$. Suppose $x_1 = 1$ and $y'_1 = 1$. Then there are two candidates for a minimal path connecting p and p', one passing through the single junction point $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$, and the other going from p to the junction point $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$, then along a straight line segment to the junction point $\left(0, \frac{1}{2}, \frac{1}{2}\right)$, and then to p' (see Figure 8.1). We can compute all the
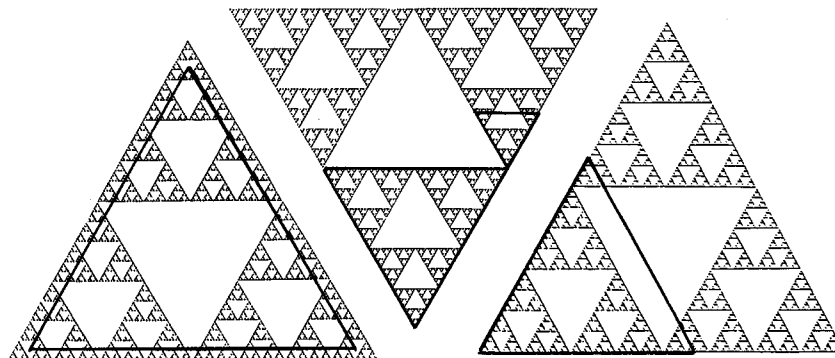
FIGURE 8.2. a) The interior triangle T_ε. b) The triangle pair T_r^2.
c) The triangular set G_r.

lengths of these portions of the curve using a scaled down version of our previous computation. For example, the minimal path joining $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ to p' has length

$$\frac{1}{2}d(X, (.x_2'x_3' \ldots, .y_2'y_3' \ldots, .z_2'z_3' \ldots)) = \left(y' - \frac{1}{2}\right)|X - Y| + z'|X - Z|$$

since $y_1' = 1$ and $z_1' = 0$. Thus we find that

(8.9)
$$d(p, p') = \min \Big\{ (x + y' - 1)|X - Y| + z|Y - Z| + z'|X - Z|,$$
$$\frac{1}{2}|X - Y| + \left(y + y' - \frac{1}{2}\right)|Y - Z| + \left(x + x' - \frac{1}{2}\right)|X - Z| \Big\},$$

and a similar formula holds in the other two cases. The minimal path is the one corresponding to the minimum in (8.9); in case the two values are equal (a line segment in p for fixed p') there are two distinct minimal paths. Considering the ambiguity in minimal paths from each vertex, we see that there are at most 4 minimal paths connecting any two given points.

Now we consider the isodiametric problem in the new metric. (In the original metric, this problem is known to be extremely difficult [AS].) We define the *isodiametric profile function* $g(D)$ to be the supremum of the measures of sets of diameter $\leq D$ (in the metric (8.1)). We restrict attention to the case of the Sierpinski gasket generated by an equilateral triangle, and for simplicity we take the side length to be 1. Of course the isodiametric estimate takes the form

(8.10)
$$g(D) \leq cD^\alpha.$$

We conjecture that the best constant in (8.10) is $c = 1$, and the extremal domains are the whole fractal or any similar image. Furthermore, for any fixed D, the extremal domain for $g(D)$ is similar to one of the two types of domains shown in Figure 8.2, the interior triangle T_ε for $D = 1 - 2\varepsilon$ or the triangle pairs T_r^2 for $D = r + 1/2$. The interior triangle T_ε consists of all points p whose distance to the two junction points (in the triangle p lies in) is at most $1/2 - \varepsilon$, so its diameter is $1 - 2\varepsilon$. The triangle pair T_r^2 is the union of one full triangle and the points of one of the neighboring triangles whose distance to the junction point is at most r, so its diameter is $r + 1/2$.

To compute the measure of these sets we introduce the function $\varphi(r)$ equal to the measure of the (triangular) set G_r of points of distance at most r from a vertex

(say X). We have explicitly

$$(8.11) \qquad \varphi\Big(\sum_{j=0}^{\infty} \frac{1}{2^{k_j}}\Big) = \sum_{j=0}^{\infty} \frac{2^j}{3^{k_j}}$$

if $1 \le k_0 < k_1 < k_2 < \cdots$, because for $r = \sum 2^{-k_j}$ the set G_r is the union of 2^j triangles contracted k_j times. We note that φ is continuous and consistent for finite sums; i.e.,

$$2^{-k_0} + 2^{-k_1} + \cdots + 2^{-k_m} = 2^{-k_0} + 2^{-k_1} + \cdots + 2^{-k_{m-1}} + 2^{-k_m-1} + 2^{-k_m-2} + \cdots$$

and

$$2^0 {\cdot} 3^{-k_0} + 2^1 \cdot 3^{-k_1} + \cdots + 2^m \cdot 3^{-k_m}$$
$$= 2^0 \cdot 3^{-k_0} + 2^1 \cdot 3^{-k_1} + \cdots + 2^{m-1} \cdot 3^{-k_{m-1}} + 2^m \cdot 3^{-k_m-1}$$
$$+ 2^{m+1} \cdot 3^{-k_m-2} + \cdots .$$

We note that φ satisfies the identities

$$(8.12) \qquad \varphi(2r) = 3\varphi(r) \quad \text{for} \quad 0 \le r \le \frac{1}{2}$$

and

$$(8.13) \qquad \varphi(r + 1/2) = 2\varphi(r) + 1/3 \quad \text{for} \quad 0 \le r \le \frac{1}{2},$$

and these essentially characterize φ (along with $\varphi(0) = 0$, $\varphi(1) = 1$). We also need the identity

$$(8.14) \qquad \varphi(r) \le r^\alpha$$

with equality if and only if $r = 2^{-k}$ for some k. This follows easily from (8.11).

Now it is obvious that

$$(8.15) \qquad \mu(T_r^2) = \varphi(r) + 1/3.$$

We claim that

$$(8.16) \qquad \mu(T_\varepsilon) = 6 \sum_{k=1}^{N_\varepsilon} \varphi(2^{-k-1} - \varepsilon),$$

where N_ε is the largest value of k such that $\varepsilon \le 2^{-k-1}$. This is clear from Figure 8.2 a), since T_ε is the finite union of 6 translates of G_r for $r = 2^{-k-1} - \varepsilon$ (T_ε is nonempty only when $\varepsilon \le 1/4$). Let D_0 be the unique value for which (8.15) and (8.16) are equal (with $D_0 = r + 1/2 = 1 - 2\varepsilon$). It is not hard to see that $7/8 < D_0 < 15/16$ and

$$(8.17) \qquad \varphi(D_0 - 7/8) = 1/145.$$

Conjecture 8.1. *For $D_0 < D \le 1$ the set T_ε (with $D = 1 - 2\varepsilon$) is the unique extremal (maximizer of measure) among sets of diameter D. For $1/2 \le D < D_0$ the set T_r^2 (with $D = r + 1/2$) is the unique extremal. For $D = D_0$, the two sets T_ε and T_r^2 are both extremal. More generally, for $2^{-k}D_0 \le D \le 2^{-k}$ and for $2^{-k-1} \le D \le 2^{-k}D_0$ the extremal sets are similar to T_ε and T_r^2.*
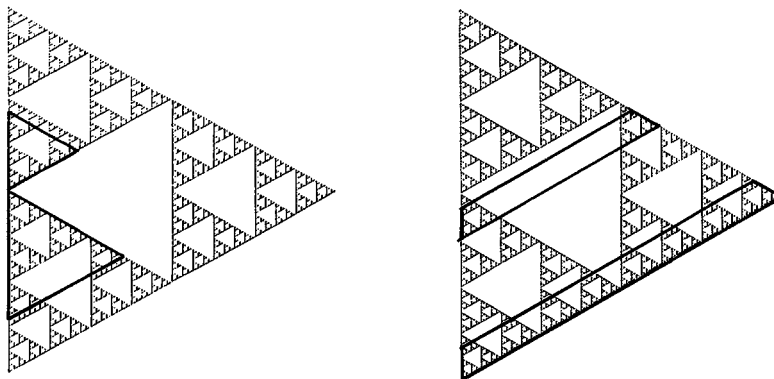
FIGURE 8.3. (a) A maximal set of diameter $r_1 + r_2$ consisting of adjacent G_{r_1} and G_{r_2} sets intersecting at a junction point. (b) Two strips of equal width r; the longer one is larger with respect to both Lebesgue measure and μ measure.

The isodiametric estimate (8.10) with $c = 1$ is an easy consequence of this conjecture. For the T_r^2 sets, we need to show that

$$(8.18) \qquad\qquad \varphi(r) + 1/3 \leq (r + 1/2)^\alpha.$$

But $\varphi(r) + 1/3 \leq \varphi(r + 1/2)$ by (8.13), with equality if and only if $r = 0$, and $\varphi(r + 1/2) \leq (r + 1/2)^\alpha$ by (8.14). Thus (8.18) holds, with equality if and only if $r = 0$. For the T_ε sets, we transform (8.16) into

$$\mu(T_\varepsilon) = 2 \sum_{k=1}^{N_\varepsilon} 3^{-k} \varphi(2^k D - (2^k - 1))$$

by using (8.12) $k + 1$ times. Since $D \leq 1$ we have $2^k D - (2^k - 1) \leq D$; hence

$$\mu(T_\varepsilon) \leq 2 \sum_{k=1}^{N_\varepsilon} 3^{-k} \varphi(D) \leq (1 - 3^{-N_\varepsilon}) D^\alpha$$

by (8.14). This shows that

$$(8.19) \qquad\qquad \mu(T_\varepsilon) \leq D^\alpha,$$

with equality if and only if $D = 1$.

We conclude by presenting some plausible evidence for the conjecture. First note that we need only consider sets of diameter D that are *maximal* in the sense that they are not properly contained in any set of diameter D. These are the analogs of sets bounded by curves of constant breadth in the Euclidean case. Now fix a value of D satisfying $1/2 \leq D \leq 1$, and suppose we have a maximal set that is contained entirely in two of the three triangles. Then it is not hard to see that it must be of the form shown in Figure 8.3 (a), the union of a G_{r_1} and a G_{r_2} set with $r_1 + r_2 = D$, joined at the junction point between the two triangles. We claim that these sets are not extremal unless r_1 or r_2 is equal to $1/2$, i.e.,

$$(8.20) \qquad\qquad \varphi(r_1) + \varphi(r_2) \leq \varphi(r_1 + r_2 - 1/2) + 1/3,$$
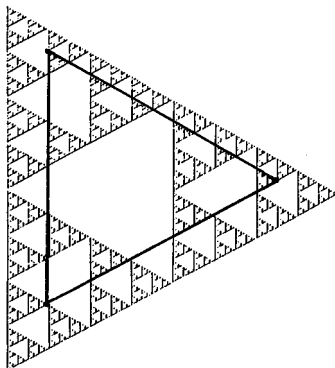
FIGURE 8.4. A maximal domain of diameter $1 - \varepsilon_1 - \varepsilon_2$.

with equality only if $r_1 = 1/2$ or $r_2 = 1/2$. But (8.20) is equivalent to

$$(8.21) \qquad \varphi(b) - \varphi(b - r) < \varphi(1) - \varphi(1 - r)$$

for $r < b < 1$, which is just the statement that the strip $G_1 \setminus G_{1-r}$ is bigger than the strip $G_b \setminus G_{b-r}$ in Figure 8.3 (b). This is certainly evident for Lebesgue measure, and can be proved for μ measure using (8.11).

For sets of diameter D that are maximal and not contained in just two triangles, the possibilities are more varied, and this seems to be an obstacle to proving the conjecture. In Figure 8.4 we illustrate one family of maximal sets, also interior triangles but not symmetric, depending on two parameters $\varepsilon_1 \leq \varepsilon_2$, of diameter $1 - \varepsilon_1 - \varepsilon_2$. When $\varepsilon_1 = \varepsilon_2$ this reduces to T_{ε_1}. (One could consider allowing three parameters $\varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3$, but such sets are not maximal unless $\varepsilon_2 = \varepsilon_3$.) It is possible to show that such domains are not extremal if $\varepsilon_1 \neq \varepsilon_2$.

REFERENCES

[AS]   E. Ayer and R. S. Strichartz, *Exact Hausdorff measure and intervals of maximum density for Cantor measures*, Trans. Amer. Math. Soc. (to appear)

[BK]   C. Bandt and T. Kuschel, *Self–similar sets 8. Average interior distance in some fractals*, Rend. Circ. Math. Palermo Sup. **28** (1992), 307–317. MR **94c:**28010

[BS]   C. Bandt and J. Stahnke, *Self–similar sets 6. Interior distance on determinatic fractals*, preprint, Greifswald 1990.

[B]   M. Barnsley, *Fractals Everywhere*, Academic Press, Boston, 1988. MR **90e:**58080

[F]   K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, New York, 1990. MR **92j:**28008

[HB]   S. Havlin and D. Ben–Avraham, *Duffision in disordered media*, Advances in Physics **36** (1987), 695–798.

[H]   J. E. Hutchinson, *Fractals and self–similarity*, Indiana Univ. Math. J. **30** (1981), 713–747. MR **82h:**49026

[K]   J. Kigami, *Harmonic calculus on p.c.f. self–similar spaces*, Trans. Amer. Math. Soc. **335** (1993), 721–755. MR **93d:**39008

[Ma1] J. Marion, *Mesure de Hausdorff d'un fractal à similitude interne*, Ann. Sc. Math. Québec **10** (1986), 51–84. MR **87h:**28009

[Ma2] J. Marion, *Mesures de Hausdorff d'ensembles fractals*, Ann. Sc. Math. Québec **11** (1987), 111–132. MR **88k:**28011

[MW]   R. D. Mauldin and S. C. Williams, *Hausdorff dimension in graph directed constructions*, Trans. Amer. Math. Soc. **309** (1988), 811–829. MR **89i:**28003

[M]   U. Mosco, *Variational metrics on self–similar fractals*, C.R. Acad. Sci. Paris Sér. I Math. **321** (1995), 715–720. MR **96h:**28013

[O]   R. Osserman, *The isoperimetric inequality*, Bull. Amer. Math. Soc. **84** (1978), 1182-1238. MR **58:**18161

[S]   R. S. Strichartz, *Piecewise linear wavelets on Sierpinski gasket type fractals*, J. Fourier Anal. Appl. **3** (1997), 387–416. CMP 97:17

Department of Mathematics, White Hall, Cornell University, Ithaca, New York 14853

E-mail address: `str@math.cornell.edu`